FORM PTO-1390 (REV 12-2001)	U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE	ATTORNEY 'S DOCKET NUMBER
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		WFU.99-35
		U S APPLICATION NO (If known, see 37 CFR 1 5 not yet assigned) 10/031889

INTERNATIONAL APPLICATION NO. PCT/US00/20400	INTERNATIONAL FILING DATE 27 July 2000	PRIORITY DATE CLAIMED 27 July 1999

TITLE OF INVENTION METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES

APPLICANT(S) FOR DO/EO/US
HART, Thomas C.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.

2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.

3. [X] This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.

4. [X] The US has been elected by the expiration of 19 months from the priority date (Article 31).

5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
    a. [ ] is attached hereto (required only if not communicated by the International Bureau).
    b. [X] has been communicated by the International Bureau.
    c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US)

6. [ ] An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
    a. [ ] is attached hereto.
    b. [ ] has been previously submitted under 35 U.S.C. 154(d)(4).

7. [X] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
    a. [X] are attached hereto (required only if not communicated by the International Bureau).
    b. [ ] have been communicated by the International Bureau.
    c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
    d. [ ] have not been made and will not be made.

8. [ ] An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).

9. [X] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).

10. [ ] An English lanuage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.

12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.

13. [X] A FIRST preliminary amendment.

14. [ ] A SECOND or SUBSEQUENT preliminary amendment.

15. [ ] A substitute specification.

16. [ ] A change of power of attorney and/or address letter.

17. [ ] A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.

18. [ ] A second copy of the published international application under 35 U.S.C. 154(d)(4).

19. [ ] A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).

20. [X] Other items or information:
Copy of Form PCT/IB/308

U.S. APPLICATION NO (if known, see 37 CFR 1 5) not yet assigned	INTERNATIONAL APPLICATION NO PCT/US00/20400	ATTORNEY'S DOCKET NUMBER WFU.99-35

21. [X] The following fees are submitted.

	CALCULATIONS PTO USE ONLY

BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):

Neither international preliminary examination fee (37 CFR 1.482)
nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO
and International Search Report not prepared by the EPO or JPO. **$1040.00**

International preliminary examination fee (37 CFR 1.482) not paid to
USPTO but International Search Report prepared by the EPO or JPO **$890.00**

International preliminary examination fee (37 CFR 1.482) not paid to USPTO
but international search fee (37 CFR 1.445(a)(2)) paid to USPTO **$740.00**

International preliminary examination fee (37 CFR 1.482) paid to USPTO
but all claims did not satisfy provisions of PCT Article 33(1)-(4) **$710.00**

International preliminary examination fee (37 CFR 1 482) paid to USPTO
and all claims satisfied provisions of PCT Article 33(1)-(4) **$100.00**

ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 100.00	

Surcharge of $130.00 for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(e))	$ 0	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	$	
Total claims	40 - 20 =	20	x $18.00	$ 360.00	
Independent claims	7 - 3 =	4	x $84.00	$ 336.00	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ $280.00	$ 0	

TOTAL OF ABOVE CALCULATIONS =	$ 796.00	
[X] Applicant claims small entity status See 37 CFR 1.27. The fees indicated above are reduced by 1/2. +	$ 398.00	
SUBTOTAL =	$ 398.00	

Processing fee of $130.00 for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(f)).	$ 0	

TOTAL NATIONAL FEE =	$ 398.00	

Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3 28, 3.31). $40.00 per property +	$ 0	

TOTAL FEES ENCLOSED =	$ 398.00	

	Amount to be refunded:	$
	charged:	$

a. [X] A check in the amount of $ __398.00__ to cover the above fees is enclosed.

b. [ ] Please charge my Deposit Account No. _____ in the amount of $ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [X] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 04-1406   A duplicate copy of this sheet is enclosed.

d. [ ] Fees are to be charged to a credit card **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO

Kathleen D. Rigaut, Ph.D., J.D.
DANN, DORFMAN, HERRELL AND SKILLMAN
1601 Market Street, Suite 720
Philadelphia, Pennsylvania 19103
U.S.A.

SIGNATURE

Kathleen D. Rigaut, Ph.D., J.D.
NAME

43,047
REGISTRATION NUMBER

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of	)
	)
Thomas C. Hart	)
	)
Serial No. 10/031,889	)
	)
Filed: January 25, 2002	)
	)
For: "Methods and Compositions	)
For Diagnosing	)
Palmoplantar Keratodermas	)
And Dysplasias and Other	)
Periodontal Diseases"	)

SUBMISSION OF SEQUENCE LISTING UNDER 37 C.F.R. §§1.821-1.825
AND PRELIMINARY AMENDMENT

　　　　This submission is in response to the April 22, 2002 Office Communication enclosing a Notification Of Missing Requirements Under 35 U.S.C. 371 In The United States Designated/Elected Office.
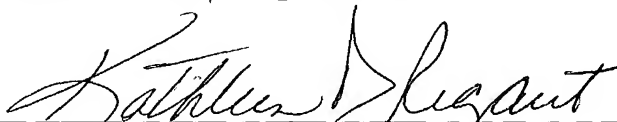
　　　　To comply with the requirements under 37 C.F.R. §§1.821-1.825, submitted herewith is a sequence listing of the nucleotide and amino acid sequences presented in the above-referenced application. The sequence listing is being submitted in both paper copy and computer-readable form. Applicants respectfully request entry of the sequence listing into the above identified patent application. The undersigned hereby verifies that the paper copy and computer readable form of the sequence listing are identical and do not contain any new matter.

In the event that a fee is required, the Commissioner is authorized to charge the account of the undersigned, Account No. 04-1406. A duplicate copy of this sheet is enclosed.

Respectfully submitted,

DANN, DORFMAN, HERRELL AND SKILLMAN
A Professional Corporation

By _Kathleen D. Rigaut_

Kathleen D. Rigaut, Ph.D., J.D.
PTO Registration No. 43,047

Telephone: (215) 563-4100

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of   )
)
Thomas C. Hart et al.   )
)
Serial No. Not yet assigned   )
)
Filed: January 25, 2002   )
)
For: "Methods and Compositions)
       For Diagnosing   )
       Palmoplantar Keratodermas)
       And Dysplasias and Other)
       Periodontal Diseases"   )

The present application is based on International Application PCT/US00/20400. Before calculation of the filing fee, please amend the above-referenced patent application as follows:

In the specification:

At page 1, line 2, please insert the following priority claim:

-- This application is a 35 U.S.C. §371 application which claims priority to PCT/US00/20400 filed July 27, 2000 which in turn claims priority to U.S. Provisional Applications, 60/145,644 and 60/165,016 filed July 27, 1999 and November 12, 1999 respectively, the disclosure of each of these applications being incorporated herein by reference.

Please amend the specification at page 80 to include a copy of the abstract which is attached hereto on a separate sheet.

In the claims:

25. The method of claim 13 wherein a germline alteration is detected by obtaining a first CTSC gene fragment from (a) CTSC gene genomic DNA isolated from said sample, (b) CTSC RNA

isolated from said sample or (c) CTSC cDNA made from mRNA isolated from said sample and a second CTSC gene fragment from a CTSC allele specific for one of said alterations, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first CTSC gene fragment and single-stranded DNA from said second CTSC gene fragment and analyzing for the presence of a mismatch in said heteroduplex, wherein no mismatch indicates the presence of said alteration.

26. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

27. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

28. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

29. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby causing a premature stop codon.

30. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a G for a A at nucleotide position 1286 in Exon 7, thereby replacing a

2

tryptophan codon with a premature stop codon.

31.   A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 1015 in Exon 7, thereby replacing a codon encoding arginine for a cysteine encoding codon.

32.   A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1019 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

33.   A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1040 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

34. A method for detecting a germline alteration in a CTSC human encoding nucleic acid, said method comprising comparing a sequence of a CTSC DNA or CTSC RNA from a human sample with an isolated wild type CTSC sequence as provided in SEQ ID NO:1.

35. A method as claimed in claim 34, wherein stability of said altered CTSC mRNA is compared with stability of wild type CTSC mRNA.

36.   A method as claimed in claim 34, further comprising expressing an altered CTSC protein from said altered CTSC encoding nucleic acid and comparing cathepsin C enzymatic activity of said altered CTSC protein to enzymatic activity of wild-type cathepsin C.

37.   A kit for detecting the presence of an altered CTSC encoding nucleic acid in a biological sample, comprising:

3

    i)   oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1;

    ii) reaction buffer; and

    iii) an instruction sheet.

    38. A kit as claimed in claim 37, wherein said oligonucleotide contains a tag.

    39. A kit for detecting the presence an altered CTSC encoding nucleic acid in a biological sample, comprising:

    i)   antibodies immunologically specific for the altered CTSC proteins of the invention;

    ii) a solid support with immobilized CTSC antigens as a positive control; and

    iii) an instruction sheet.

    40. A kit as claimed in claim 39, wherein said antibody contains a tag.

REMARKS
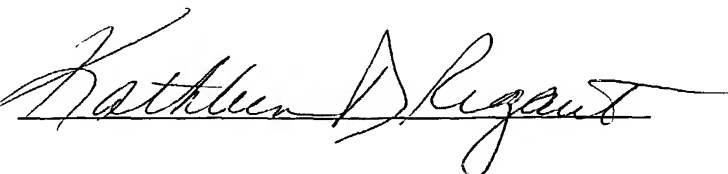
    The purpose of this preliminary amendment is to 1) correct a minor error in the numbering of the claims; 2) insert a priority claim into the specification and 3) provide a copy of the abstract on a separate sheet.

    Favorable consideration leading to prompt allowance of the present application is respectfully requested.

            Respectfully submitted,
            DANN, DORFMAN, HERRELL AND SKILLMAN
            A Professional Corporation

            By _____

4

Kathleen D. Rigaut, Ph.D., J.D.
PTO Registration No. 43,047

Telephone: (215) 563-4100

Enclosures: Marked up copy of the claims showing where
amendments have been made
        Copy of an abstract of the specification on a
separate sheet

Marked up draft of claims:

[26] <u>25</u>. The method of claim 13 wherein a germline alteration is detected by obtaining a first CTSC gene fragment from (a) CTSC gene genomic DNA isolated from said sample, (b) CTSC RNA isolated from said sample or (c) CTSC cDNA made from mRNA isolated from said sample and a second CTSC gene fragment from a CTSC allele specific for one of said alterations, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first CTSC gene fragment and single-stranded DNA from said second CTSC gene fragment and analyzing for the presence of a mismatch in said heteroduplex, wherein no mismatch indicates the presence of said alteration.

[27] <u>26</u>. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

[28] <u>27</u>. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

[29] <u>28</u>. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

[30] <u>29</u>. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby

6

causing a premature stop codon.

[31] <u>30</u>. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a G for a A at nucleotide position 1286 in Exon 7, thereby replacing a tryptophan codon with a premature stop codon.

[32] <u>31</u>. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 1015 in Exon 7, thereby replacing a codon encoding arginine for a cysteine encoding codon.

[33] <u>32</u>. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1019 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

[34] <u>33</u>. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1040 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

[35] <u>34</u>. A method for detecting a germline alteration in a CTSC human encoding nucleic acid, said method comprising comparing a sequence of a CTSC DNA or CTSC RNA from a human sample with an isolated wild type CTSC sequence as provided in SEQ ID NO:1.

[36] <u>35</u>. A method as claimed in claim 3[5]<u>4</u>, wherein stability of said altered CTSC mRNA is compared with stability of wild type CTSC mRNA.

[37] <u>36</u>. A method as claimed in claim 3[5]<u>4</u>, further comprising expressing an altered CTSC protein from said altered CTSC encoding nucleic acid and comparing cathepsin C

7

enzymatic activity of said altered CTSC protein to enzymatic activity of wild-type cathepsin C.

[38] 37. A kit for detecting the presence of an altered CTSC encoding nucleic acid in a biological sample, comprising:
    i) oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1;
    ii) reaction buffer; and
    iii) an instruction sheet.

[39] 38. A kit as claimed in claim 3[8]7, wherein said oligonucleotide contains a tag.

[40] 39. A kit for detecting the presence an altered CTSC encoding nucleic acid in a biological sample, comprising:
    i) antibodies immunologically specific for the altered CTSC proteins of the invention;
    ii) a solid support with immobilized CTSC antigens as a positive control; and
    iii) an instruction sheet.

[41] 40. A kit as claimed in claim [40]39, wherein said antibody contains a tag.

ABSTRACT

Compositions relating to mutations in the CTSC gene associated with palmoplantar keradermas are disclosed. Also provided are methods for using the compositions of the invention in diagnostic and prognostic screening assays.

9

METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES

FIELD OF THE INVENTION

The present invention relates to the fields of
genetic screening and molecular biology. More
specifically, the invention provides compositions and
methods that may be used to advantage to isolate and
detect a palmoplantar keratoderma predisposing gene,
cathepsin C (CTSC), some mutant alleles of which cause
susceptibility to certain pathological disorders, in
particular Papillon-LeFevre Syndrome, Haim-Munk Syndrome
and certain forms of early onset periodontal diseases.
More specifically, the invention relates to germline
mutations and functional polymorphisms in the CTSC gene
and their use in the diagnosis of predisposition to
palmoplantar ectodermal disorders/ dysplasias and
periodontal diseases. The invention also relates to the
therapy of palmoplantar ectodermal disorders/dysplasias
and periodontal diseases which have a mutation or
functional polymorphisms in the CTSC gene, including
gene therapy, protein replacement therapy and protein
mimetics. The invention further relates to the screening
of drugs for treating and alleviating disease symptoms.
Finally, the invention relates to the screening of the
CTSC gene for disease-related mutations, which are
useful for diagnosing the predisposition to additional
disorders and dysplasias, including but not limited to
prepubertal periodontitis, early onset periodontal
disease or other forms of gum disease.

BACKGROUND OF THE INVENTION

Various publications or patents may be referenced in this application by numerals in parentheses to describe the state of the art to which the invention pertains. Full citations for these references are provided at the end of the specification. Each of these publications or patents is incorporated by reference herein.

Most forms of inflammatory periodontal disease can be successfully treated and managed. As a result, the ultimate goal of periodontal therapy has changed from that of simply arresting disease progression to one aimed at regenerating the supporting tissues. Unfortunately, not all forms of periodontal disease respond to treatment. Severe periodontitis that is resistant to conventional periodontal treatment has been recognized in a number of monogenic conditions. Certainly some of the most intriguing and dentally challenging of these conditions include Papillon-Lefevre syndrome (PLS), Haim-Munk syndrome (HMS) and periodontal diseases.

In 1924, Papillon and Lefevre described two siblings, the products of a first cousin mating, with a condition characterized by diffuse transgradient palmoplantar keratosis (PPK) and the premature loss of both the decidous and permanent dentitions. This condition came to be known as Papillon-Lefevre syndrome and subsequently over 200 cases have been described. The hallmarks of PLS are palmoplantar keratosis and rapid periodontal destruction of both dentitions. An increased susceptibility to infection has been reported in approximately 20% of PLS patients. Additional findings include intracranial calcifications, retardation of the somatic development, follicular hyperkeratosis and onychogryphosis. Clinical findings reported in PLS patients

suggest that the clinical expression of the condition is

2

highly variable. Unfortunately, to date, no
pathognomonic disease marker exists allowing definitive
diagnosis of PLS.

In 1965, Haim and Munk described an unusual

5   syndrome in four siblings of a Jewish religious isolate
from Cochin, India [21]. In addition to congenital
palmoplantar keratosis and progressive early onset
periodontal destruction, other clinical findings shared
by these individuals included recurrent pyogenic skin

10   infections, acroosteolysis, atrophic changes of the
nails, arachnodactyly, and a peculiar radiographic
deformity of the fingers consisting of tapered pointed
phalangeal ends and a clawlike volar curve. Subsequently
pes planus was reported to be associated with the

15   syndrome [24]. This was the first reported association
of these clinical findings, and the condition became
known as Haim Munk syndrome, or keratosis palmoplantaris
with periodontopathia and onychogryposis (HMS;
MIM245010)[22]. Although the palmoplantar findings and

20   severe periodontitis were suggestive of the
Papillon-Lefevre syndrome (PLS; MIM245000)[3], the
association of other clinical features, particularly
nail deformities and arachnodactyly, argued that HMS was
a distinct disorder.

25   PLS and HMS are classified as type IV palmoplantar
ectodermal keratodermas [2]. The unique presence of
severe, early onset periodontitis distinguishes PLS and
HMS from other PPKs and raises the question of whether
they result from the variable clinical expression of a

30   common gene mutation, are allelic mutations at the same
genetic locus, or result from expression of gene
mutations at separate loci. Although Haim and Munk's
initial report proposed HMS was a distinct entity,
Hacham-Zadeh and co-workers referred to the disorder as

35   Papillon-Lefevre syndrome and cited Gorlin's suggestion
that HMS was a clinical variant of PLS [23,25]. In his
review of PLS cases reported in the literature, Haenke

[5] summarizes an extensive list of clinical findings reported in PLS affected individuals, including increased susceptibility to infections, ectopic cranial calcifications and nail anomalies [5,26]. It is unclear if these additional clinical features are coincidental findings that may be segregating in a particular family or if they are etiologically related to a syndrome with a very variable clinical expression. Because PLS is an uncommon condition, and generally occurs only in a single generation it is difficult to determine if these occasional reports of associated clinical findings are etiologically related to PLS. Additionally, consanguinity is common among parents of PLS cases and therefore, it may be expected that an increased number of rare recessive conditions may be seen. Such is likely the case for the reports of mental retardation associated with PLS [5].

Pre-pubertal periodontitis (PPP) is a rare and rapidly progressive disease that results in destruction of the periodontal support of the primary dentition. The condition may be localized (usually to deciduous molars) or generalized. The localized form begins at approximately 4 years of age and is associated with only mild gingival inflammation in the presence of relatively little plaque. The generalized form begins earlier, immediately after eruption of the deciduous teeth. It is associated with severe gingival inflammation and hyperplasia, although significant gingival recession has also been described as an associated clinical feature. The attachment loss appears to be continuous rather than intermittent as with most other forms of periodontitis.

A varied clinical phenotype has been reported for PPP, probably reflecting the fact that the term PP describes an etiologically heterogeneous group of conditions that share an overlapping clinical presentation. Although PPP can occur as an isolated finding, many reports of PPP describe an increased

4

systemic susceptibility to infections. Children with PPP
have frequent inner ear infections and infections of the
respiratory tract [39,40]. Prepubertal periodontitis is
known to be associated with Papillon Lefevre syndrome
5   and with a number of systemic disease states that share
an increased susceptibility to microbial infections.

    To date, no pathognomonic disease marker exists for
most PPKs allowing for definitive diagnosis. The
present invention provides such a disease marker and
10   methods of use thereof having diagnostic and prognostic
utilities for several PPKs and many periodontal
diseases.

SUMMARY OF THE INVENTION

15   The present invention provides compositions and
methods which allow for genetic screening and diagnosis
of certain palmoplantar keratodermas and periodontal
disease states in affected individuals. In accordance
with the present invention, it has been discovered that
20   mutations or functional polymorphisms in the cathepsin C
gene (CTSC) give rise to certain pathological conditions
including PLS, HMS and periodontal diseases. Mutations
or functional polymorphisms associated with the disease
state are those which give rise to a altered, truncated,
25   misfolded or otherwise non-functional CTSC polypeptides.
Polymorphisms in the CTSC sequence which do not affect
the nature of the encoded protein are not associated
with PLS, HMS or periodontal disease.

    Thus, in one embodiment of the invention, a method
30   is provided for determining the presence of alterations
in CTSC encoding nucleic acids which give rise to
altered CTSC proteins. The wild-type CTSC nucleic acid
sequence and its encoded amino acid sequence are known.
See SEQ ID NOS: 1-3 provided herein. This sequence
35   information facilitates the identification of genetic
changes that give rise to aberrant CTSC proteins.

    CTSC mutations specifically associated with PLS,

5

HMS and PPP are described herein and are set forth in Table 1. Accordingly in one embodiment of the invention, nucleic acid molecules encoding altered CTSC proteins are considered to be within the scope of the

5   present invention. In a preferred embodiment of the invention, the altered CTSC nucleic acid has at least one of the alterations set forth in Table 1.

       In a further embodiment of the invention, nucleic acid probes which specifically hybridize to the human

10   altered CTSC-encoding nucleic acids described herein and not to wild-type CTSC encoding nucleic acids are provided. In a preferred embodiment, the probes specifically hybridize with altered CTSC encoding nucleic acids having at least one of the alterations set

15   forth in Table 1.

       In yet another embodiment of the invention, a mutated CTSC protein encoded by the altered CTSC encoding nucleic acids of the invention is provided. Preferably such CTSC proteins are encoded by a nucleic

20   acid containing a mutation as set forth in Table 1. Also provided are assays for biochemically assessing altered cathepsin C activity. Antibodies immunologically specific for altered CTSC proteins are also contemplated to be within the scope of the present invention.

25   In another aspect of the invention, a method for detecting a germline alteration in a CTSC gene is provided. In a preferred embodiment the alteration is selected from the group consisting of the alterations set forth in Table 1. The method comprises analyzing a

30   sequence of a CTSC gene or CTSC RNA from a human sample or analyzing a sequence of CTSC cDNA made from mRNA from a human sample and comparing sequences so isolated to the wild type sequence encoding CTSC. Inasmuch as certain alterations of the CTSC coding sequence may not

35   alter the function of CTSC, methods are provided for assessing the enzymatic activity of proteins encoded by nucleic acid molecules which do not possess the wild

6

type CTSC sequence.

In yet another embodiment of the invention, kits are provided for detecting the presence of an altered CTSC encoding nucleic acids in a biological sample. An exemplary kit comprises the following: i) oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1; ii) reaction buffer; and iii) an instruction sheet. Kits for detecting the presence an altered CTSC proteins in a biological sample are also provided. Exemplary kits for this purpose comprise: i) antibodies immunologically specific for the altered CTSC proteins of the invention;   ii) a solid support with immobilized CTSC antigens as a positive control; and iii) an instruction sheet. Optionally, anti-CTSC antibodies used for this purpose may contain a detectable label or tag for used in isolating or detecting immune complexes.

Various terms relating to the biological molecules and cells of the present invention are used throughout the specifications and claims.

With reference to nucleic acids used in the invention, the term "isolated nucleic acid" is sometimes employed. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule. An isolated nucleic acid molecule inserted into a vector is also sometimes referred to herein as a Arecombinant@ nucleic acid molecule.

With respect to RNA molecules, the term "isolated

'7

nucleic acid" primarily refers to an RNA molecule
encoded by an isolated DNA molecule as defined above.
Alternatively, the term may refer to an RNA molecule
that has been sufficiently separated from RNA molecules
with which it would be associated in its natural state
(i.e., in cells or tissues), such that it exists in a
"substantially pure" form (the term "substantially pure"
is defined below).

With respect to protein, the term "isolated
protein" or "isolated and purified protein" is sometimes
used herein. This term refers primarily to a protein
produced by expression of an isolated nucleic acid
molecule of the invention. Alternatively, this term may
refer to a protein which has been sufficiently separated
from other proteins with which it would naturally be
associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a
preparation comprising at least 50-60% by weight the
compound of interest (e.g., nucleic acid,
oligonucleotide, protein, etc.). More preferably, the
preparation comprises at least 75% by weight, and most
preferably 90-99% by weight, the compound of interest.
Purity is measured by methods appropriate for the
compound of interest (e.g. chromatographic methods,
agarose or polyacrylamide gel electrophoresis, HPLC
analysis, and the like).

With respect to antibodies, the term
"immunologically specific" refers to antibodies that
bind to one or more epitopes of a protein of interest,
but which do not substantially recognize and bind other
molecules in a sample containing a mixed population of
antigenic biological molecules.

With respect to single stranded nucleic acids,
particularly oligonucleotides, the term "specifically
hybridizing" refers to the association between two
single-stranded nucleotide molecules of sufficiently
complementary sequence to permit such hybridization

under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary

5   sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence. Appropriate conditions enabling specific hybridization

10   of single stranded nucleic acid molecules of varying complementarity are well known in the art.

For instance, one common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a

15   specified sequence homology is set forth below (Sambrook et al., 1989):

$T_m = 81.5°C + 16.6Log [Na+] + 0.41(\% G+C) - 0.63 (\% formamide) - 600/\#bp$ in duplex

20

As an illustration of the above formula, using [Na+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1%

25   decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C.

The term "promoter region" refers to the transcriptional regulatory regions of a gene, which may

30   be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. In the present invention, the use of SV40, TK, Albumin, SP6, T7 gene promoters, among others, is contemplated. Specific promoters for the yeast and mammalian expression systems

35   of the invention are available and known to those of ordinary skill in the art.

The term "operably linked" means that the

9

regulatory sequences necessary for expression of the
coding sequence are placed in the DNA molecule in the
appropriate positions relative to the coding sequence so
as to enable expression of the coding sequence. This
5   same definition is sometimes applied to the arrangement
of transcription units and other transcription control
elements (e.g. enhancers) in an expression vector.

The phrase "functional polymorphism" refers to a
change in wild type CTSC coding sequence giving rise to
10   altered cathepsin C activity as assayed using
conventional methods.

The term "oligonucleotide," as used herein refers
to primers and probes of the present invention, and is
defined as a nucleic acid molecule comprised of two or
15   more ribo- or deoxyribonucleotides, preferably more than
three. The exact size of the oligonucleotide will
depend on various factors and on the particular
application and use of the oligonucleotide.

The term "probe" as used herein refers to an
20   oligonucleotide, polynucleotide or nucleic acid, either
RNA or DNA, whether occurring naturally as in a purified
restriction enzyme digest or produced synthetically,
which is capable of annealing with or specifically
hybridizing to a nucleic acid with sequences
25   complementary to the probe. A probe may be either
single-stranded or double-stranded. The exact length of
the probe will depend upon many factors, including
temperature, source of probe and use of the method. For
example, for diagnostic applications, depending on the
30   complexity of the target sequence, the oligonucleotide
probe typically contains 15-25 or more nucleotides,
although it may contain fewer nucleotides. The probes
herein are selected to be complementary to different
strands of a particular target nucleic acid sequence.
35   This means that the probes must be sufficiently
complementary so as to be able to "specifically
hybridize" or anneal with their respective target

10

strands under a set of pre-determined conditions.
Therefore, the probe sequence need not reflect the exact
complementary sequence of the target. For example, a
non-complementary nucleotide fragment may be attached to
the 5' or 3' end of the probe, with the remainder of the
probe sequence being complementary to the target strand.
Alternatively, non-complementary bases or longer
sequences can be interspersed into the probe, provided
that the probe sequence has sufficient complementarity
with the sequence of the target nucleic acid to anneal
therewith specifically.

The term "primer" as used herein refers to an
oligonucleotide, either RNA or DNA, either
single-stranded or double-stranded, either derived from
a biological system, generated by restriction enzyme
digestion, or produced synthetically which, when placed
in the proper environment, is able to functionally act
as an initiator of template-dependent nucleic acid
synthesis. When presented with an appropriate nucleic
acid template, suitable nucleoside triphosphate
precursors of nucleic acids, a polymerase enzyme,
suitable cofactors and conditions such as a suitable
temperature and pH, the primer may be extended at its 3'
terminus by the addition of nucleotides by the action of
a polymerase or similar activity to yield an primer
extension product. The primer may vary in length
depending on the particular conditions and requirement
of the application. For example, in diagnostic
applications, the oligonucleotide primer is typically
15-25 or more nucleotides in length. The primer must be
of sufficient complementarity to the desired template to
prime the synthesis of the desired extension product,
that is, to be able anneal with the desired template
strand in a manner sufficient to provide the 3' hydroxyl
moiety of the primer in appropriate juxtaposition for
use in the initiation of synthesis by a polymerase or
similar enzyme. It is not required that the primer

11

sequence represent an exact complement of the desired template. For example, a non-complementary nucleotide sequence may be attached to the 5' end of an otherwise complementary primer. Alternatively, non-complementary

5   bases may be interspersed within the oligonucleotide primer sequence, provided that the primer sequence has sufficient complementarity with the sequence of the desired template strand to functionally provide a template-primer complex for the synthesis of the

10   extension product.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1E are a series of clinical photographs showing palmoplantar keratosis and periodontal disease

15   in PLS study patient. Fig. 1A: palmar hyperkeratotic lesions; Fig. 1B: plantar hyperkeratotic lesions; Fig. 1C: hyperkeratotic lesions affecting the knees; Fig. 1D: periodontitis involving erupting permanent dentition and Fig. 1E: periapical radiographs showing severe alveolar

20   bone loss affecting erupting permanent teeth.

Figure 2 shows haplotype data for chromosome 11q short tandem repeat polymorphisms (STRP) markers spanning the PLS gene locus. Segments which are likely

25   to be homozygous by descent are boxed. Arrows indicate recombinant events. Individuals 7 and 22 share a common haplotype for D11S1979, D11S1887, D11S1780, D11S1367, D11S931, and D11S4175.

30   Figure 3 depicts pedigree and sequence analysis of *CTSC* exon 6. The numbering of the wildtype sequence shown above the figure is based upon the genomic sequence of *CTSC*. See SEQ ID NO: 1. Circles represent females and squares represent males. Filled symbols

35   indicate affected individuals. Half-shading indicates carriers based upon DNA sequencing results. All affected individuals are homozygous for the specific CTSC

mutations. Arrows indicate the position of the mutation. This family has a nonsense mutation (856 C->T) at codon 286 resulting in a truncated protein of 286 amino acids.

5   Figures 4A - 4D show pedigrees and sequence analysis of *CTSC* exon 7 for 4 Families with PLS. Symbols are as described for Figure 3. Fig. 4A: Family with a single base pair deletion (1047delA) of *CTSC* resulting in a frameshift and premature termination.
10   Fig. 4B: Family with a 2bp deletion (1028-1029delCT) of *CTSC* resulting in a frameshift and premature termination. Fig. 4C. and Fig. 4D: pedigrees of families with a nonsense mutation (1286G->A) at codon 429 resulting in a truncated protein of 428 amino acids. The
15   father in family C is deceased and no sample was available for analysis.

          Figures 5A-5C is a schematic diagram of the CTSC gene showing the locations of the mutations described
20   herein. Panel 5A. Genomic structure of *CTSC* gene with introns shown as solid lines and exons depicted as boxes. The 5' and 3' untranslated regions are shown as filled boxes. Panel 5B. Coding region of *CTSC* gene. The amino acid numbers are shown at the end of each exon.
25   Mutations listed in Table 1 are shown according to their genomic locations with Missense, *; Nonsense, ◆; Insertion, ■; and Deletion, □. The splicing site mutation is indicated by an arrow. Panel 5C. Subunit structure of *CTSC* polypeptide with SP, signal peptide;
30   P1, 13.5 kDa pro-region; P2, 10 kDa pro-region; H, heavy chain; and L, light chain. The 10 kDa pro-region is cleaved out upon activation. The disulfide bond within the 13.5 pro-region is shown. The glycosylation sites are indicated by filled circle and arrows indicate the
35   active sites.

          Figures 6A-6F show a series of micrographs

depicting the clinical and radiographic findings in Haim
Munk syndrome. Fig. 6A: dermal involvement of fingers in
individual #34. Fig. 6B: Individual #34 radiograph of
terminal phalanges of the fingers showing marked

thinning increasing towards the distal, tapering pointed
ends showing a claw-like volar bend. Fig. 6C: Individual
#17 palmar keratosis; Fig. 6D: Individual #17 plantar
keratosis: Fig. 6E: Individual #17 gingival
inflammation; Fig. 6F: Individual #17 radiograph showing

alveolar bone destruction associated with gingival
inflammation shown in 6D.

Figure 7A shows pedigree of Cochin descendents
segregating Haim Munk syndrome (HMS). Numbered
individuals have been analyzed for the current study.
Circles = females, squares = males, shaded symbols = HMS
affected individuals. Double lines = consanguinity.
Individuals #10,11 * = second cousins. Numbers inside
circles, squares and diamonds indicate the number of
additional offspring not examined in this study.
Sibships described in previous reports are indicated and
referenced below the pedigrees. The subjects of Haim
and Munks original report (1965) are individuals 33, 34,
35, and 36. Half-shading indicates carriers based upon
DNA sequencing and/or restriction enzyme analysis.
Unshaded numbered individuals represent non-carriers
based upon DNA analysis. Figure 7B shows a pedigree of
a Turkish family segregating PLS. Numbered individuals
were available for study. Half-shading indicates
carrier based upon DNA analysis. Individual 77 is a
non-carrier based upon DNA sequencing.

Figures 8A and 8B show the results of sequence
analysis of exon 6 of CTSC. The numbering of the
wildtype sequence is based upon the cDNA sequence of
CTSC. See SEQ ID NO: 1. Fig. 8A: Family A (Cochin
isolate diagnosed with Haim Munk syndrome) from Figure

1. Affected individuals are homozygous for a 857A->G missense mutation which results in a conserved glutamine being changed to an arginine (Q286R). Representative sequences are shown for individuals #36 (affected) and #31(carrier). Fig. 8B. Family B from Figure 1. Affected individuals are homozygous for a 856C->T nonsense mutation which results in a premature stop codon at position 286 (Q286X). The Q286X mutation has been previously reported in an inbred Turkish family [12].

Figure 9 depicts a gel showing the results of restriction enzyme analysis of Q286R mutation in Family A of Fig. 7. A 465 bp fragment of exon 6 was amplified and subjected to restriction digestion with AvaI as described under methods. The Q286R mutation introduces a new AvaI site. After digestion and electrophoresis through 1.8% agarose gels, wildtype individuals exhibit bands of 465 bp, affected individuals have bands of 404 and 61 bp, and carriers have bands of 465, 404, and 61 bp. M. 1 kb ladder (Gibco). Lane 1. Individual #5 uncut, demonstrating 465bp amplicon. Lane 2. Individual #5 cut with AvaI. Only the 465 bp fragment is observed. Thus individual #5 has the wildtype sequence on both alleles. Lane 3. Individual #31 uncut. Lane 4. Individual # 31 cut with AvaI. The 465 and 404bp fragments are visible, confirming that individual #31 is a carrier of the Q286R, consistent with the sequencing results shown in Figure 3A. Lane 5. Individual #34 uncut. Lane 6. Individual #34 cut with AvaI. The 404 and 61bp fragments are indicated by arrows.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect mutated forms of the lysosomal

15

protease cathepsin C (CTSC) gene, associated with
autosomal recessive disorders characterized by palmar
hyperkeratosis and/or periodontitis. More specifically,
the present invention relates to germline mutations in
5   the CTSC gene and their use in the diagnosis of
predisposition to such pathological conditions.
Additionally, the invention relates to germline
mutations in the CTSC gene in other palmoplantar
ectodermal disorders and dysplasias and their use in the
10   diagnosis and prognosis of such pathological conditions.
The invention also relates to the therapy of
palmoplantar ectodermal disorders and dysplasias which
have a mutation or functional polymorphism in the CTSC
gene, including gene therapy, protein replacement
15   therapy and protein mimetics. The invention further
relates to the screening of drugs which may have
therapeutic value. Biochemical assays are provided for
the assessment of altered activity of aberrant CTSC
enzymes encoded by the mutated CTSC encoding nucleic
20   acids of the invention. Finally, the invention relates
to the screening of the CTSC gene for mutations, which
are useful for diagnosing the predisposition to
ectodermal disorders and dysplasias.
        The present invention provides an isolated
25   polynucleotide comprising all, or a portion of the CTSC
locus or of a mutated CTSC locus, preferably at least
eight bases and not more than about 100 kb in length.
Such polynucleotides may be antisense polynucleotides.
The present invention also provides a recombinant
30   construct comprising such an isolated polynucleotide,
for example, a recombinant construct suitable for
expression in a transformed host cell.
        Also provided by the present invention are methods
of detecting a polynucleotide comprising a portion of
35   the CTSC locus or its expression product in an analyte.
Such methods may further comprise the step of amplifying
the portion of the CTSC locus, and may further include a

step of providing a set of polynucleotides which are
primers for amplification of said portion of the CTSC
locus. The method is useful for either diagnosis of the
predisposition to PPKs or the diagnosis or prognosis of

5   keratodermal disorders/dysplasias and periodontal
diseases.

The present invention also provides isolated
antibodies, preferably monoclonal antibodies, which
specifically bind to an isolated polypeptide comprised

10   of at least five amino acid residues encoded by the
altered CTSC locus.

The present invention also provides kits for
detecting in an analyte a polynucleotide comprising a
portion of the CTSC locus, the kits comprising a

15   polynucleotide complementary to the portion of the CTSC
locus packaged in a suitable container, and instructions
for its use.

The present invention further provides methods of
preparing a polynucleotide comprising polymerizing

20   nucleotides to yield a sequence comprised of at least
eight consecutive nucleotides of the CTSC locus; and
methods of preparing a polypeptide comprising
polymerizing amino acids to yield a sequence comprising
at least five amino acids encoded within the CTSC locus.

25   The present invention further provides methods of
screening the CTSC gene to identify mutations. Such
methods may further comprise the step of amplifying a
portion of the CTSC locus, and may further include a
step of providing a set of polynucleotides which are

30   primers for amplification of said portion of the CTSC
locus. Exemplary primers are set forth in Table A.

The method is useful for identifying mutations for
use in either diagnosis of the predisposition to
keratodermal disorders/dysplasias and periodontal

35   diseases or the diagnosis of such disorders.

The present invention further provides methods of
screening suspected CTSC mutant alleles and functional

polymorphisms to identify mutations in the CTSC gene. In addition, the present invention provides methods of screening drugs for therapy and to identify suitable drugs for restoring CTSC gene product function.

5   Finally, the present invention provides the means necessary for production of gene-based therapies directed at aberrant cells associated with keratodermal disorders and dysplasias. These therapeutic agents may take the form of polynucleotides comprising all or a

10   portion of the CTSC locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the CTSC protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire

15   protein sequence of CTSC. These may functionally replace the activity of CTSC in vivo.

It is a discovery of the present invention that mutations in the CTSC locus in the germline are indicative of a predisposition to keratodermal

20   disorders/dysplasias and periodontal diseases. The mutational events of the CTSC locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

A major gene locus associated with the keratodermal

25   disorders and dysplasias of the invention has been localized to a 2.8 cM interval on chromosome 11q14 of the human genome. This region contains a genetic locus, CTSC. The CTSC message is expressed at high levels in a variety of immune cells including polymorphonuclear

30   leukocytes, macrophages and their precursors. This gene is expressed in the palms, soles, knees, and oral keratinized gingiva.

The CTSC gene was originally reported to consist of 2 exons. US Provisional Application 60/165,016 from

35   which the present application claims priority, describes mutations in Exons 1 and 2. The mutations described actually fall within Exons 6 and 7. Reference numerals

to the altered amino acids are the same as those in US
Provisional 60/165,016, only the nucleotide numbering
has changed to reflect the actual genomic structure of
the CTSC gene which is now known to contain 7 exons.

5   The sequence encoding the wild type human CTSC gene is
provided below (SEQ ID NO: 1):

```
AGGGAGATAT AAGTGAATAA TTTGGACCTG CTCTCTTTGA ATGTTTATAA TCTGGTGGAA
AAAAAATGGA CATATGAATA TTGATTTGTG ACCAGTGCAA AGGGGGCAAA AATTCATATC
CCAAAGAAAA CGGGGACACA TCAGGTCTGT CTTGTTCATC ACTGTGTCCA CAGGGCCTGA
CACCTAGTAG GCTCAGTGGG AGAAAGGAGC CCCAATTACC AACAAAAGCC AGGAAAGAAC
GGGAGGCTCT TACGGAAAAG GGTGATACTT AAACTGAGCA AGGAGGCACC TGGAAATAGT
GCCACCTAAT AATTTTTGGC GATCAGACTG ACACACTAGA ACGGTTCATA AGACCAGCCT
TCTCCCATTG GCTAGCTTCC TTCCTCACCC TTCTCACCCT GGGCAAGCCG CTTCCTCTCT
CTGGGCCTCT TGCTTTTCCT CTGTAACATA AAAGGGGTTG AGCAATATCA TCTCTGAGAG
CGCCATGTGT GTGCGTGCCA GAGGGAAAAC CCCCACAACG CTAATACATC AAAACTGCAG
GTTTGCACAA AAACTGAATT CTGCTGAATG CAAACAGGCA AACAGCATTT ACCAGGAAAC
AAAACAAAAT CAAGCACATA AAAAGTAGG AAGAGTTGGA AAACGGAAGG AAGATAAGTT
CTCAAACAGC TGGAATAGTT GATGTTAGCT AGCGAAGTTT TTCAGAGGAA AAAACAAGAA
GTTGGTTATG AGGCAAGTGG ACCTGAGAAA AAAGACTAAA GGGGAAGAAT AGCAAGTAAA
ACAGAACTCC ACTTGCTAGA TCTCTCCCTC TGTCGCGCTC TTTCACCTGA CCCACTCCCT
TATTCCCCCC ACACCCTTTC CTTCTCTCCC TACGTTACCG CACAGGAACG AAGTCTGGGT
CATGTGCGGA CCGCTTGTGG CTCTTAAATC CTCTTTTTGT CACCCTGGCC GTGCAAAATT
TTGAAACGTC CCTCGGCAAA AAAAATAAAA ATAAAAAAAA AAAATCTGTC CCTGGCCTCT
TCCCTAGTTC TGGGTCCAGT TGCAGCCAAG TGAGGGGCAG CGCGCGCTCC CAAGTCCCCG
TTTCAGAGAC GCGCACGCGC CTGGCGCCCA ACCCCCAATC CCCTGCTGCT CAGTGACCCC
GCCCACGGGT TTCCGGGCCG GCGTAGCTAT TTCAAGGCGC GCGCCTCGTG GTGGACTCAC
CGCTAGCCCG CAGCGCTCGG CTTCCTGGTA ATTCTTCACC TCTTTTCTCA GCTCCCTGCA
GCATGGGTGC TGGGCCCTCC TTGCTGCTCG CCGCCCTCCT GCTGCTTCTC TCCGGCGACG
GCGCCGTGCG CTGCCGACACA CCTGCCAACT GCACCTATCT TGACCTGCTG GGCACCTGGG
TCTTCCAGGT GGGCTCCAGC GGTTCCCAGC GCGATGTCAA CTGCTCGGTT ATGGGTAAGC
CGCCGGCTCG GCAGTCCTCC GGGTCGTCCT TTCTGCCCTT GAGCCCCTAA CGCAGCGCCA
CGCCAACTAC CGCTTCCCCC CAGGCAGACG CTTGTGGGTG GCCAGAGCAT CTTGACTGGA
TTCGGGGACC CTTGGGGACC TTCTTCCCCG CCAGGCTCGC GAAGTTAAAG TTCATCTGCT
GAGAACTTCT AACTCCACAC TTTCTTGGTT ATCTTGGGGA CTCAACACTT TGATCAAGAA
CTTTTTTATT CCTCCCGCTT AATTTTGTTT GCTTTGAGAG AGACTTGGGA ACTGCAATCG
TTTGGTTCTC CAGTCCGATC TGGTAGCGTT ATTTTTAAAA TTTATTTTTA TTTTTTATTA
CTATTTTACT AGTGAAGATA GATGAGCTCA GAGACTCTCG AGGATATAGC ATGAAGTTTT
CTCTTTTTGT TAGATGGTGG GAAAGGGACT TTCTGCCCAG CGATTTTGGT TTGAGCGGGT
GTTGATGAGT ACTAGAAAAC GGCTAGTACC ACTCTGCATT GTTTCATGCA TTGCAAGGAG
GTAAAAATTT TTTAAAAAAT TAATAACAAA GAAAACTTAA CTCTGAACCT AGTAATTAGA
AATGCCCAGA GTCTGCACAA TGTTTGGCTC ATGGAAGGCT CTCAATAAAT ACCTAGTGTT
TGAACATACT GGAGATATTC CATATGCCTT CAGATAACAT GGTTACCCCT AGAACAAAGA
```

```
ACCTAGTGAA  GGGGGTGGGG  TGGAAAAAAG  ACAATCACTA  GTTAGAAGAG  TCACACTGTG
GTCTACCCAA  ACCCTCTTAC  AGCCTGTGAC  TTCTGCAGAG  TCAGTAAAAA  ATCAGCTATA
ATTTTCTTGT  CAACAGAACA  AAGATTAAGT  CATTTGTTAC  TAAAGAAATA  CCTTTTTAAC
TGACTGTATA  GCATTTCATA  ATCCTGAACA  TTGTAACTTT  TTTTTTTTTT  TTTTTGAGAA
TGGAGTCTTG  CTCTGTCACC  CAGGCTGTAG  TTCAGTGGCG  GGATCTCGGC  TCACTGCAAC
CTCCGCCTCC  CGGGTTTAAG  CGATTCTCCT  GCCTCAGTCT  CCCGAGTAGC  TAGGATTACG
GGCACATGCC  ACCACGCCCG  GCTAATATTT  TGTATTTTTA  GTAGAGACGG  GTTTCACTGT
GTTAGCCAGG  ATGGTCTCCA  TCTCCTGACC  TCGTGATCCG  CCCGCCTCGG  CCTCCCAAAG
TACTGGGATT  ACAGGCGGGA  GTCACCACGC  CTGGCCTTTT  TTTTTTTTTT  TTTTTTAAAT
TTGAGTTTGA  AGGTTGGGGC  AGAAAGAGAT  CAGGATTTGC  ACTGCCCTGT  CACATGCAAT
CTCCCATGTC  AGAGCATTAC  TCTCAAACAT  GGAAAAACTT  TAAAATACAC  AACTCTCCAG
ATGGACACAG  CTGAATCATT  TTCAGGAAGG  CTCTGCCTAT  AAAATTACCT  TTGGCCTTAA
TTCAGTAATT  AAAGGCACCC  ACAGGTCTGA  GCCCTCATCT  ATGAAGGTTA  AACGTATCAT
TCCTATCAGC  ACTAATTGGT  TTTATAGGAT  ACAGACCTCT  AGTTTGCTGA  ATAAACTTTG
GAAGGATTCA  GATTATGAGG  TTCTAAACTG  TAGAGCCTTT  GAGGAGAAAG  GTACCCATT
TCTTCTCCGA  ATAGATCATT  TTGTGTCTTC  TCCTGGGCTT  AGCACATTGT  CTTCCTTAGG
ATCTGATAGT  CTGTTTATTT  ATCTTTTTGT  TGTCATACCT  TTTATTCTTG  CATTTCCCAC
TTTTGACTAG  CATTTTGCCT  TTTCTCCGTT  TCTGGAAGCC  TGTAATTTTC  AACATTCCCC
ATTTCTCCTT  TTGTTCAGTG  GAAAAATTTC  TTCAGTGTTG  GGATGCTCTG  GAAAGGTACT
GTAGTTTTGG  GGGTCTCCCT  GCCTGCTGGG  CTGCATAGCA  TATCTCTCTT  TTTTGAGACA
AGGTCTCACT  GTGTTGCCCA  GGTTGGAATG  CAGTTGTGCA  ATCATAGCTC  ACTGCAGCCT
CCAACTCCTG  GGCTCAAGCA  ATCCTTCCAC  CTCAGCTTCA  GCCGCTCAAA  TAGCTGGGAC
TACAGGCAAT  GCACCACCAT  AACCAGCGAA  TTTTTAAAAA  TTATTTTTGT  AGAGACAGGG
TCTCACTATG  TTGCCCAAAC  TAGTCTTGAA  GTTCTGACCT  CAGCCCCGCA  AAGTGCTAGG
ATTACAGGCG  TGAGCCACTG  CTTCTGCTTG  CATTATCTCA  AATTTTTAGA  GCCTAACTTC
ACAGTCTTGC  CTCGGAGCAG  GAACTGTTTG  AGGCAACAAG  ATGGAGCCTC  TGGTTTCTTC
ACTAGGCAGA  CTGTGCTCAA  ACTGGGTAGC  ATGAAAGGAA  AGCAGCACAA  TTAAAATTGA
AATTGGGGGA  TTCTTTCCCC  TATTGAAATC  CAAATAATTT  TCCTTTATTG  TGCTTTTTTT
TTTTTTCTTT  AGGACCACAA  GAAAAAAAAG  TAGTGGTGTA  CCTTCAGAAG  CTGGATACAG
CATATGATGA  CCTTGGCAAT  TCTGGCCATT  TCACCATCAT  TTACAACCAA  GGCTTTGAGA
TTGTGTTGAA  TGACTACAAG  TGGTTTGCCT  TTTTTAAGGT  TAGTTTTGTT  GGAAGTTGGA
TTTACATTTT  CAATGATTTG  ATATCTGAAA  CCTCTTCTGA  TTAGTAGACC  CTCAGAATTT
TAATTTTAGA  TTAAGAAGAT  GACCGGAATT  GACACCACTC  TTCCCAAGAG  TAGTAGTAGT
TGGTATAATT  TTGCCACTTT  ATTTAAAATT  ATGATTTTAG  TAGGTTTACA  AAATACCAGT
GCTACATTTG  AATATGTATA  AATTATGTTT  AAAATTTACA  TTTTGGTAAG  TATGACTAAA
TTCTTAATTT  ATTTTCCTTA  TTACTACCAC  TTTATTTCTA  AATGTTGCCA  TAGTCATTTG
GCTTTGTTCT  AAATCTGTAG  GAAAGATAGA  GAGATTACAC  ATTTTGTTTT  CTTGCAGTTA
CTATGCTGTC  CTTCCTATCA  CTACCTGTTG  GCTGAGGTAG  TGATAGGCCT  AAATGATTCA
TTATCTTAAA  TGTACTAAAT  ATGTTGAGTA  ATTTTTTCTT  CTAAACTAAC  AGAAAGAGAG
AACCTAGGAG  TTACTCCCTT  AGGCTGGTTA  AAGTGAAAGG  TAGCCAAGTC  AACCCAGCTT
GTTTCCTTCT  CTCATTAGGA  AAGAACTATT  GTTCATTCTC  ATAACACACT  TTTTCCAATT
GCAAACATAC  TCAGGGTTAA  AATAGTTTAG  CACAAATTGC  AGCCCATTTC  ATTTGTTCTT
CACAAGCTGG  AACTTTTCTT  GTAAGCTAAA  TATTAAATGG  TTCAAGTAAA  TTGGATACAT
AAGCCTGAAA  CTAGGCGTTT  CTCATTATAC  ATAGAGTATA  AATTAAGACA  GACTTTTTCA
```

```
     TGGTGAAAGG TTTACAGCCT TTAAAACATC TGGGAAGAAG TGGGAAAGTA GGGAATAACT
     CTGTTAAATA TGATAAAAGA CAAAGCACCA ACAAAGGCCT AGTTCTAAAC TTGTTATAAT
     TTCTCATGGG AGTTGTGGTT TGTCACAAGG TTATGGCGGT CCAAGCAAGT TTACAATATT
     TTTTAGAATA ATAACTCCCA GAAATATTTT TAAAATAAGG ACCTTTCTTT AATATGGAAA
5    AAAAAAAGAT GAAATAGAGA GGAAGAGGTT GCCTTTCCTC AACTGGATTT GCTGGTGTGA
     TATGCAGCTG GTGGAAACCA GTTTGTCCAT GCTAGCTGTT GATGGCCACC AGTCCACTGA
     ATTAGTGGAA ACCCATGCTT AACCAGTTGT TAACTATTTT TAGTATCATT TCCAGATGCT
     ACGCATTTTT TTTGTAAAAA ACATACTTAC AATAGTATCT ATATAAATTT TTAAAATTGT
     AATGTATATG TGGTATATTT AACCTGAAAA TAATCTTTGG TGTATACGTT GATGAGCCTA
10   GGCCTTTGCA GACCTCTTAC AGTTACTCTG TCACAGCCCT TCAAAGGTTC TTTGACTTCA
     GATTAAGAAT CAATTGCATG TGGAATGCAT GTGCAAAGGA AGAGATATTC AAGGCAATTG
     TTACCATGTA CCATAAACCT TGTACATAAT TTTTTGTCAT TTTCTTTTCC TCTGTCTCTA
     TCCCTTCTTC TTTGACACAC AATAGACACC TAGTCAATTT ATTACAAAAA AAATGAATGA
     ATGAAGTGGA ATTCAGTTGG GAAATAGGTT AAATAAATTA TTTAGGAGAT GAGGAATAGG
15   TAAAAGAGA CAAGTACATA GTTTATTCTT TTGACTTAGA AAACTTTTGA TTCTTAAATT
     CTGCAGAATT GGAGAAACTG GTGGGGAAAC TTCTAAAATC ATTATTTAAT TACCAGAGAT
     GTAATAGATA TAGACAAAAG CAGTTTTCTT CTTTTATTAT TTTTTCATCA GTTAGTTCTT
     AGCTTAAATA GTAGTCCAAA GCTGGTAGGG ACAGAGGGAA TTAGCTGGTG GCTGAATGAG
     GAATTGTATC ACTTTTTGTG AATCACGGTG TAAGCACATT TGGTGTTTTG CCATTGCCTA
20   AGAACATTAG TCACATTAGG TCAATAGAAA ATCACTTTTT AAAGCCAAAT AAAGTTATAT
     GTGTTCCCAA CCATGAGTTG GAAAGAATTA ATATATATGC TGTTGGAGGG TAGAACCCTG
     CCTAATCATA TGGTTCTGGA TGGCATTGAT CGAATCCTTA TTCTTTCATT AGGAATAACA
     ATAGAAAAA TACTCCTGCC CTACTGATTT CAGGATATGT CTATTTAAA GTGCCCATTT
     GACAAAACCA TTATCAGGGC CATGTTTTCT TTTTCTGCAG AAAAATCAAC CACTCTGGTC
25   AGTAGTTAGG TCTTATGACA AGCACCATAA TTTCCTTAGG CAGAGTAGAA TATAATAGGA
     TACTTCTTTT TGAAACTTAA TATAATCAGG TAGTTCCAGA TAAACATAGC TTGCAAAGTG
     ATAAAATACC ATGTTATTTT AGTAAATCCA ATTGCAAGAG TGATGGGAAA CAGAGTTTAA
     AAACTTAAGA AAGATATTAA GATGGAGTTG ACTTTGAATA ATAAAGTCAT CCACTGTTGA
     TGGGTGACAT TTATTAATAC AGAAAGTTTA CAGATTTTAC CATAAGCATC AGGGTATTTC
30   CTGCAGCTGG GGAAACCGTG CTTGAAAGGA TGCGTAACTC AAGGAAAACA CAAGCCCATG
     TAATAAGTAT TGCATGTGAG AATTGTTCCT ATAGAATTAG AAAGCATCTT TTACATTAAA
     ATTTATTTTT GTAAAAGAG AAAACATCAA AACTTGAGTA GTATTTGCTA TTCAAAGAGT
     CTTACATAAA CGAAAACATA CTCAACCTAC TGCCATTACA GAAATATTTG ACAAATTCTT
     GCCATGCTTA CCTGCCATCG TTGTTGTTAT CCTACAAATC AATTGGATTT TCACGCCTCT
35   CCACTGACTG GAACCCTACA ACTTGCTTCC TTTTATCCTC TTTATATATG CTTCAGATAT
     GCTTGAAGTA GATTGTTTCT TATTGTTCTT GCTGCTCGAG TTTGTTGAGT AGTTGGTACA
     TGCAGAGTAT TTGTATGTTA TGACATATAT AGGTTTTGTC CATGGTTCCT GGCTCATAAC
     TCACTCCCAA GACCCTTGTT ACAGAAACCA GAATCTCTCT CTCTGATCTT CTCCTACCCT
     CCTTTCATCT GCCCACTGCA GAACTCTAAT CTGATTATGG TTTCTAAGAC CCTCATACCA
40   GAGAGTATTC TGCCCCATAC CATAGCAGAA GGAACACTGC ACAGAGACAC CAAGAAGAAT
     CTGAACAGAC AGGCCTTGTT AGGTTTAGAT CATGTCCTTA TAACCTAATT ATATTTTAAC
     ATGGTTATCC ATGCTTTAAT CATGTGTATT CAATGAACGC TCCATAAAAG CCCAAGAAGA
     ACAGATTTGA GGGAATTCTG AAGCGCTAAA CATGTAGGGT CTGACAGGAA GGTGAAGAAG
     AACTCATCAT GCTGGAAGGG TGGTCCACCC TAACTCCGCA GGGACAGAAG CTCCTGTGTC
```

```
        TGGGACCCTT CCAGAACTTG CTCTATGTAT CTCTTCAAGT GCCTGTGTAT TTATATCCTT
        TCAAATATCC TTTGTAATAA ATCATAAACA TGTTTCCCTG AGTTTTTATG CCACTCTGTC
        AAATTAATTG AACCCAAAGA GGGGGTTATG GGAACCCCAA TTTGAAGCCA GTCCATCAGA
        AGTTCTGGAG GCCTGGACTT GAGACTGGTG TCTAAAGTGG GCAGGCAGTC TTGGGGACTA
  5     AGCCCTCAAC CTACGGGATC TGAAACTGTC TCCTGGTAGG TAGCATTGGA GTTGTACTGG
        AGGGCACTGA GCTGGTGTCT GCTGCAGAAT TGATTGCTTG CTTGCTGGTG GGGAGAACTT
        CCTACATATT TTGGGGTCAC CGAAGTCTTT TGTATTGATT GTTGTTGCTG AGCTCCTATT
        GCTAGAAACT TGCTTTATTA AGTTCTGTTT ATTCAGGGCT TATCTGAAAG AGAAACATTT
        TTTATGATTT GAGATTTCTA AGCCATTTTA AAACTTGGTT TTAACAGCTT GAGAAATTGG
 10     GGGTATAGTA GGGATGGAGA TACATATATT TGGATGTGAC TTCAAGCTAG ATAAAAGTTT
        GGAAGGAATA AAAGTTTGAC ATCTAAAGTT TTTGCATAGT TTGAGTTGAG CAGGAGGTAC
        TAGGTATGTT TTTAAAATAT TTTTTTCAGC CAGGCACGGT GACTCATACC TGTAATCCCA
        GCACTTTGGG AGGCCGAGAC AGGCGATCAC CTGAGGTCAG GCGTTCGAGA CCAGTCTGGC
        CAACATGAAG AAACCCCGTC TCTACTAAAA ATACAAAAT TAGCTGGGTG CGGTGACACA
 15     TGCCTGTAAT CCCAGCTATT CAGGAGGCTG AGGCAGGAGA ATTGCTTGAA CCCAGGAGGC
        AGAGGTTGCA GTGAGCCGAG ATCACACCAT TGCACTCCAG CCTGGGTGTC TCAAAATAAT
        AATAATAATA AAAAAATGAA AGATTTTTTC TTACTCAGCA TCCTCCAGGC ATTTTATTAT
        CTGAGCACTT TATGGGAGTT GCATATTACA TTTAGGGCCC ACTCAGGTGG GTGGGTATCT
        AAGCATTTGA AATAACCTTA TGTAAACTAA TAAGGAGTAA TCAGGCCTGT GGCAAGATGG
 20     AAACAGTCTT AGAGGCATTC AAATTCAAAT TTCCTTTAAA ACACTGGGCT GGCCAAAACA
        AAAGACAATA CTATCTACAG GCCAGTTTCT AAGATTATCA GATTTTAGTA GCATTTACCA
        TTTCATTGTA CTTGGCACAC TTTAGCAAAT TTGCACTTCT TAAAAGTACC TGCAGGCAAT
        CTCCTATATA AAAACACAAT GCAGGCTAGC TTGGCTCCTG CCTTTAATTC CAGCACTTTG
        AGAAGTTGGA GACTAGCGTG GCCAACGTGG TGAAACCTCA TCTCCACTAA AAATACAAAA
 25     ATTAGCCAGG CATAGCAGCG CACGCCTGTA GTCCCAGTTA CTTGGGAGGC CGAGGCAGTA
        GAATCACTTG AACGCTGGAG ACAGAGCATG GAGTGAGCTG AGATTGCACC ACTGCACTCC
        ATCCTGGGTG ACAGCGTGAG ACTCTGTCTC CAAACAAAAC AAAACACACA CACACACAAT
        GTAACAACAC GAAACAGAAT ACTGTGAAAA TGCTTAATTA TGTCTGACTT TACATGATGG
        CTGGATATGT GATTATTTTT TCTTCTTTAT GCTCTTATGT ACTTTGTTCA TTTTTAATGA
 30     TGATCATGTA TAAAGCTCCT CTGTGTAGCA TTCTCTCCAC CAAATTGCCC AGAGACAGGA
        AGTCCTGTAA AACAAACTAA GCTCCAAAAA ATGACCTCCT GTTGAATAGG CTTTTTTTTT
        TTTTTTTTTT TTTTTTTTTG AAATGGAGTC TAGCTCTGTC TCCCAGGCCC TCGCTCCTTC
        CACCTCCTGG GTTTAAGAGA TTGTCCTTCC TCAGCCTCCA GAGTAGATTG GATTACAGGT
        GCCCGCCATC ACGCCCAGCT AATTTTTGTA TTTTTAGTAG AGATGGGGTT TCACCATGTT
 35     GACCAGACTG GTCTTGAACT CCTGACCCCA AGTGATCCGC CCGCCTGGAC CTCCCCAAGT
        GCTGGGATTA CAGGTGTGAG CCACCACGCC CAGCCTGAAT AGGCTTTCTA ACCTACTGTT
        TCTCATTTTA CTTTCTCTGA GGCAGTAAAA GAAACTGACT CTAAAAGGGA GCAGTAGAGA
        AGGAACTCAG ATTTTATTTT GAAGATTAAG CTACTCAAGG GCTGAGGAAA TATGTAGAGG
        GGAAGTAGAC TCATTATGGC TGGAAATTTT ATTTGGTGAT AGTGAAGCAA ATCTTAGGGC
 40     TTTCTAATTG AGCTCTTGTT TGAAAGGCTC CAATCTTAAT AGAACTATAA GCTAAAAAAA
        ATGACCATCA GTGTTTCTAA AGCAAGTTGC TACTCAAAC AAGAACACTT TGGGAGGCTG
        GGGCTGGTGG ATCACCTGAG GTCAGAAGAT GGAGTCCAGC CTCAACATGG TGAAACCCCT
        CTATACTAAA AATACAAAAA GTAGCCGGGT GTGGTGGTGC ACGCCTATAG TCTCAGCCAC
        CTGGGAGGCT GGGGCAGAAG AATCGCTTGA ACCCGGGAGG TGGAGGTTGC AGTGAGCTGA
```

22

```
      CATTGTGCCA CTGCATTTCA GCCTGGGTGA CAGAGTGAGA CTCTGTCTCA AAAACAAAAC
      AAACAGAAAC AAGAACTCTA TGTTGAGAAA TCCATACTAG AGGGTTTAAT TTCTCACTTT
      GTGTGCTAGT GATTAATAAA TTTCAAGCTT ATCACACAGC CAGAAATGTC GCCTGTGTTT
      CTACAATAAA AGATTTGGGA ATATTGTGAC ATTTTTTCAC TGAGCCTTCT GGGTTCATTA
5     TTAAGATATA ACAATTTTAG AAGACTTATT GAGATAGGTA TACTTTTTAA AATTTGGATT
      CAATATATCA AGCTCAGCAT TTTGTCTTTT TTTTGTTTTG TTTTAAACAA CTTTGGGTTT
      ACTTGTAAGA GTATTTCAGT GGAGGAAAAG TATCAGAAGT TTAGCACTGA GTGCCTGCCT
      AACTATTTTA TGTCTTCTTT CATGATGATG CTTAATCTCA ACATAATGAC TCAGTTTACC
      ATTGTAAACT CTTTGTAGGA TGTCACTGAT TTTATCAGTC ATTTGTTCAT GCAGCTGGGA
10    ACTGTGGGGA TATATGATTT GCCACATCTG AGGAACAAAC TGGGTGAGCT CCATTGAAAA
      ATTTCTAACT TTCTTATTAA AAGCAAGCTT CCTCACTTCC TGATGGCTTA TGCAAATTAA
      TGCCTATTTA TTCATCAGAG GGTGACTGGA GAGATTATGG TTTTCTTATG AATTTCTTGC
      TTGAGCCTGC TTGTCTGCTG TTTCTAAGTT GTCTGCACTA TTTTATGTGA GTAATTTTCT
      ACTTTATTAT GTTCTGTTTT CACATGTCAA ATCAGCTCTC CCAAGAATGC TACTTGTAAC
15    CTAAGTAGAC GTGAACCGAA AAGGGTAAAG ACCCAGCTAA AAAAAAGGTT GACTCAAGTT
      CAGTTCACAT TCGTAAAGTA GTTAGGCTGC CTTAGGTTGC TTATTCTTTT CTTGAAAAGA
      ATCCCTCAAG AGAGAAACAT GTGAGGCCAC AGCAGCTTAG ATCTGTCTTC CACAGAGAAG
      GTGGCTTTTA CAGAGAAATT GACATACTCA TCACTTATCT GACATGACCC AGCTTTGTAA
      AACTGGCTTC TATTAAAATA GTCTTAATAG ACTATTCACT GAGGAGGAGA GAATCTTATT
20    CACTCTTTAC ATTCTCTTCA CATTTTCAGA ATAGATGTTT AAATCATTGC TCACACTGGA
      TCCATAAATG TCTAAAATGT TGATGAAGAA ATAGGTGATT AGAGAGTAAA ATTAATAGAG
      ACTTACCCTT CCTTGCATTT TAACATAATA TTCTTTCCCC TTTTCCTTCC TCTGTTACTT
      GGCTCTTAAA TACCAGAAGT GAGATATGAA AAAGGGAACT GGGAACAAGT ATTGAAAGCA
      CCATAGGTTT ATCTTATATT AGCATTTTCC AAACTTTATA ATGAACCAGC AGTGACCCAC
25    ACATCTGCCA GGTAGGAATT GCTCATCAGT TGTGTCGCGT TATCTTGTTA AGCTTCAGCA
      ATGTTCCATA CAGCCACTAA TAACAGATCA AAGTGAGCAT TAGGGTTGAA ATTAGTAAGC
      TGTTTCTTCT CAGTTCTTTC TGGTAGTTGA ATAAATATAG AATGTATTAA ATAGTTTTCT
      TTATTTCAGA ACTTCTCAGA GTCTGTAATA TATTGTATGG TGGTAGCTTA GGAGGAAGAT
      GCAATAGGAA ACTTTTCCCA GATAGGTTCA CTATTTTTTT TTTCCACGAA AAATAAGCTG
30    TTCTCAAAAT ACAGTTTACA AAATTTTATC CTTAACTCTT CACTCTTTCT CCTAGTTAGG
      GAGACCGCTC CACCAGTAGA AAAGATAAAC CCTGGTAATT TGTTGTGTAA ATGGGATAAA
      TAGCCTAGTA CCTAGTCATG TGGATTCAGG CAGCACTGAG CCTAAATTAA AGTTTGCAAG
      GTATACATGT TAATGTATCT AAGTTACTAT ATTTAGCCTG TTTCTTAAGT ATGTTTCAGA
      AACATATTCG TTTTTTTCAG TGGCAGTTAC CTTCAGATGC ATGTGCTTCT AAAGCATGTT
35    GGTTGCATGT GCAGTACATG TTGCTTAAGC ATTTAGCTTC AGAATGGCAT CTTTCCTGTG
      AATGTCTTAA CATTTACAAA AATATACCAG GATCTCAAAT ATCAGTGCTG GTATTTTTTT
      TTTTTTTTAC TTAAAGAAAC TGATATGATT AAATATTAAG AGACAATATG ATCCTTGTTG
      GCTTGTAACC CTAGTTTTTA TTGTCTTGTA GTTATTAAAT AGAGCATCTG TTGAGGGACT
      CTTTTAAAAC CACAGCCATG AACAGACGTT GGGGCTAAGA GACAGAGCAG CCTGCGACAG
40    TGTGGACCTA CCTGTAGCAG CTAGCAAAGG CCTCTAGCAG CTACAGTCCC TTCTGGAGTC
      TTTATTTGCA TGCAAAATGC AAAGGAGTCC TGGTGACCTA CCTCCAAGGC AGCTGCCCTC
      CTGAACACTC CCTTGGAAAA CAGTAAACAT CATTTTGGAA TGTGAACAAC CAGAGACTAC
      ACAGGAGAAA GGAAAAAAAA ATTCTGAAGA TGCAAAATCT TGGGTGGCTT CACCGTTCAG
      TTTTTTAATA AAAGGAACAA TATACAACAC GTTGTTCTTT TTCTCTTTTG AAATCCCTTC
```

```
      TATTACAGTG ATTTTTTTCT AAGATTGTCA GGATTTGAAG TGTATGTTTT GTTTTATTCA
      CAGCGTAAAT TTTATTCACA GTTTAACTGT CTGCCTGAGT GTCTTTCCTT TCTCTAATTA
      CCTTGAGGAA CCCAAGAGCC TGTTGTAAGG AGAAAATAAG GCCCTTGGAT CTCTTGAGAT
      TCACAGATAT AAGTTATTGA AGGGAAGATG GTCCTATGGA GGACATATTT AAAGAAGGGA
  5   AAAAAGAGGC TTTCTCAGAT ATGTCAGACT GCTATAGTAT ACTTCTACAG ATTATAGACC
      TCCAGTACCT CTGGCCAGAA AGATGGTATC GTAAACACCC TATTTTTTTT CTTTTCTTTT
      TTCATTAGGT ACAAGCTTTG TGCTAAGAAG TTGACATACT ATAAGCTACA AAAGTTCTGT
      AAAGTAGATA TAACTAGTTT CATTTTATAG ATAGAGAAAA TTAATCTCTT ACAGTGCTAA
      GCTCACAGAG TTTCTAACTG TAAAATGCTA GAACTTGTCT TTCAAGCCTA AAGACTTCCT
 10   TGGGGCTAAA TAGTGAAAAA AGCCATTTCA CAAATAAGTA AATGGTATTT AGAGGCATAT
      TTGGATTTCC TGGTAAATTC CAGTCTGTGA GCATCATGAA TATTAGTTTA ATGTTGCATG
      GGCTCATGTT GAAGTTTTAA GAGAAGAACT GCCTTGAAGC TTAGGTTTCC TTAGCTATTA
      GGCTACTGAC TTTCTTGCCT AAACCAGGGT TTTTTCATTG AAGACCAAAA CTTACCTTCT
      CCTTCAGTTT GTAGTTTGGA AATTGGTAGA AGAGCTTTGT AAACTTCAAA TTAAGTACAA
 15   ACTAAGTGTC ATAGTCAAAT TTACTAATCT TAATTACAGT ATTGTTCAAC TGATTGCTAT
      CTTCTAGCTC TTTCCTGCCG AATAATGGTC TTGTTTCCTG CTCTGTTGGT TTAGAGCTGA
      CTTCTTTCAG CTTTGGTAAG CCTGAAATTA TGGGGTTATG TTTAATTCAT ATTGTCTGGG
      TGGACTTTCC TCTCTTGCAT TTCTGCTTGA ATAGAAGAAT TTTTCTCTAG AGAGTAGTTT
      GTCATCCTTA CTCTGTTGAT TCAGATGACT CTTTGTATGA TCTGAGAGGT ATACTGTTCT
 20   GCTATTCTGA GAAGAAGTAT TTCAGAAAGA TGAATTAAGA GTACAGTGGA CTGCTCCCAC
      CTGGAAACTT TTATCTATCT CACCTCTGGA CCTGATAAAT TCTTTATCAC TCAGGACCTT
      GATGACGCTG CTCTCTGAAA CCCTCCCCAG CTCTCTCTAT TACCGTGAGA AACATCAGAA
      CTTTGGTTCC CATTGCATAT CGCAGGTACC TCTGCTTTCA TGCCATGCTG TAATGGAGTG
      ATTGGGTAGC ATGTTTTCAT CTCTTTCCAG ATTGAAAATC TGTATTTCTC CCTGTATATC
 25   TTCAACACCT AATGCACATA GAACTTTGTA GGTACCTGGA AAATGCACCA CAGTTTTCTT
      TTCTTTTTGC AGACTTTTCA CAAGTATTAC CAACTTACAA AGAATTAATT TTGTAGGATT
      CTAGAAAGAC AAATCAGGAA  TGGTGCCATA TACATCTTTT TTGATTCCCT GCTCTAAAGA
      ATATTATCAG GTTACCTTCC TGCAGAGTTT TAAAAGAATT GCATATTTCA AGCTGACTTT
      CAGGATGTAA ATATAACCAA AGCAACTGAT ATGTAAAAAA TATATTCAAT GGCATTCCTA
 30   GATTTTCTTC TAGGGTGTTT TATTGTTTTG GGTTTTACAT TTAAGTCTTT AATCCATCTT
      GAGTTAATTT TTGTATAGGT ATAAGAAAGG GGTCCAGTTT TAATTTTCTG CGTATGGCTA
      GCCAGTTCTC CCAGCACCAT TTATTAAATA GGGAATCCTT TCCCTATTGT TTGTTTTTGT
      ACGGTTGTC AAAGATTAGA TGGTTGTAGA TGTGTGGTCT TATTTCTGAG ATCTTCATTC
      TCTTCCACTG GTCTATGTGT CTGTTTTTGT ACCATGCTTT TTTGGTTACT GTAGCCTTGT
 35   AGTATAGTAT GAAAGATAGC ATGATGCCTC CAGGTTTGTT CTTTTTGCTT AGGATTGTCT
      TGGCTATACG AGCTTTTTTT TGGTTCTATA TGAATTTTAA AATAGTTTCT TCTAATTGTG
      TGAAGAATGT TAATGGTAGT TTAATGGGAA TAGCATTGAA TCTGTGAATT GCTTTGGGCA
      GTATGGCCAT TTTCATGATA TTGATTCTTC CTATCCATGA GCATGTAACG TTTTTCCCTT
      CGTTTGTGTC CTCTCTCATT TCCTTGAGTA GTGGTTTGTA GTTCTCCTTG AAGAGATCCT
 40   TCACTTCTTC TGTATTCCTA GATATTTTAT TCTCTCTGTA GCTATTGGGA ATGGGAGTTC
      ATTCATGATT TTGCTCTCTG CTTGCCTTTT GTTGGTGTAT AGGGATCCTG GTGACTTCTG
      CACATTGATT TTGTATCCTG AGACTTTACC GAAGTTGCTT ATCAGCTTAA GAAGCTTTTG
      GGCTGAGATG ATGGGGTTTT CTAGATATAG GATCATGTTA TCTTCAAACA AAGACAATTT
      GACTTCCTCT CTTCCTATTT GAGTACGCTT TATTTCTTTC TCTTGCCTGA TTGCCCTGGC
```

24

```
        CAGAACTCCC AATACTATAT TGAATAAGAA TGGTGAGAGA GGGCATCCTT GTCTTGTGCC
        AGTTTTCACG GGGAATGCTT CCAGCTTTTG CCCATTCAGT ATGATATTAT CTGTGGGTTT
        CTCATAAAAA GCTCTTATTA TTTGAGATAC GTTCCTTCAA TACCTAGTTT ATTGAGAGTT
        TTTAACATGA AGCGATGTTG AATTGTATCG AAGGCCTTTT CTGTGTCTAT TGAGATAATC
   5    ATGTGGTTTT TGTCTTTAGT TCTGTTTATG TGATGAATGA CGTTTATTGA TTTGCATATG
        TTGAACCGGC CTTGCATCCT GGGGATGAAG CCAACTTGAC TGTGGTAGAT AAGCTTTTGG
        ATGTGCTGCT GGATTTGGTT TATCAGTATT TCATTGAGAT TTTTTGCGTC GAAGTTCATC
        AGGGATATTG GACTGAAGTT TTCTTTTTGT TGTCGTATCT CTGCCAGGTT TTGGTATCAG
        GATGATGCTG GCCTCATAAA ATGAGTTAGG GAGGAGTCCC TCCTTTTCAA TTGTTTGGAA
  10    TAGTTTCAGA AGAAAGGGTA TCAGCTCCTC TTTGTACCTC TGGTAGAATT CAACTGTAAA
        TCCATCTGGT CCTGGACTTT TTTTCATTAG TAGGCTATTT ATTACTGCCT CACTTTCATA
        ACTTGTTATT GATCTATTCA GGGATCCAAC TTCTTCCTGA TTCAGTCTTG GGAGTGTGTA
        TGCATCCAGG AATTTATCCA TTTCTTCTAG ATTTTCTAGT TTCTTTGCAT AGAGGTGTTT
        GTAGTATTTG CTGTTGGTTG TTTGTACTTC TGTGAGATCA GTGGTGGTAT CCTGTTTATC
  15    ATTTTTTATT GTGTCTGTTT GATTCTTCTC TTATTTTTGA CAAAGCTGAC AAAAAGAAGC
        AATAGGGAAA GGACTCTCTA TTCAATTAAT CCTACTGTAT ATCTGGCTAG CCATATGCAG
        AAAATTGAAA CTGTTCCTGT TTCTTAATCC ATATACGAAA ATCAACTTAC GATGGATTAA
        AGACTTAAAT GTAAAACCCA AAATTATAAA ACCCTGGAAT AGAATATAGG CAATATCATT
        CTGGACATAG GAATGGGCAA AGATTTTATG AGAAAGACAC CAAAAGCAAT TACAACAAAA
  20    GCAAAAATTG GCAAATGAGA TCTAATTAAA CTAAAGAGCT CTGCACAGCA AAAGAAACTA
        CTGTCAGAGT GAACAGGCAA CCAACAGAAT GGGAGAAAAT TTTTTCAATC TATCCATATG
        ACAAAGGTCT AACATCCAGA ATCTACAAGG AACTTAACAA ATTTACAAGA AAAAAGGAGC
        CCCATTAAAA AGTTGGCAAA GAACATGAAC AGACACTTCC CAGAAGATAT TCATGTGGCC
        AATAAACATG AAGAAAAGCT CAACATCACT GACCATTAGA GACGTGCATA TCAAAATCAC
  25    AATGAGATAC CATCTCATGT CACAATGGTG ATTATTAAAA AGTCAAACAA CATGCTAGTG
        AGGTTGTAGA GAAATAAGAA CGCTTTTACA CTGTTGGTGG GAATGTCAAC TAATTCAACC
        ACTGTGGAAG ACAGTGTGGT GATTCCTCAA GGATTTAGAA CCAGAAATAT CATTACTGCA
        TATAGACCCA AAGGAATAGA AATCATTCTA TTACAAAGAT ACATGCACAT GTATGTTTAT
        TACAGCACTA TTCACAATAG CAAAGACATG GAATCAACCC AAATGCTCAT CAGTGATAGA
  30    CTGGAAAAAG AGAATGTGGA ACATAAACAC CATGGAATAC TATGCAGCAA TAAAAAGGAA
        TGAGATCCTG TCCTTTTCAG GGACATGGAT GGAGTTGGAA GCTGTTATCC TCAGCAAACT
        AATGCAGGAA CAGAAAACCA ACCACCACAT GTTCTCACTT ATAAGTGGGA GCTGAACAAT
        AGAACACATG GGCACAGGGA GGGGAATAAC ACACACTGGG GCCAGTCAGG GGGTGGGGGG
        TCAAGCTGAG GGAGAGCATT AGAAAAAATA GCTAATGCAT TCTGGGCTTA ACCCATTTAT
  35    GCCTAGTGTT CCATTTCTGG AATGCTAAGC ATGTGGAAGT TCTTTATATC CTGCTCAAGG
        TCATTGCCAA GGTCTGATTT TTCACATTCA ACAAATTGCA ACCTCTGGCA TAAATGGGTT
        AATACCTAGG TGATGAGTTG ATAGGTGCAG GAAACCACCA TGGCACATGT TTATCTATGT
        AAGAAACCTG CACATCCTAC ACATGTACCC TGGAACTTAA AAAATTTAAA ATATATATGT
        ATATATATTT AATATGGAAT TTTAAAAATT ACTAATGAGT CTTTTATCT GAGTAATTTT
  40    GCATCAACAT GCTTTTATTA TGGAAGAGAA GATTCAGTGA GTACAAAATT GCAGATACAT
        GTGTCAGAAG ATCCCTGAAT ATAATAAGGC TTAGTATTCT GTGTCATAAT TGCCTGTTTG
        TATTCCTCTC TGGTCTTTAA ACTTCATTAG GGCAAGGATC AACTCCATCT TACTAACCAT
        TTGATTCCCT ATGTATTACA CGATATATGA CCAATAATAA GCCTTCAATA AATACTTGTA
        AAATAAAGAA TGTTATGTAA TATATCATGT GGTATTGTTT TATTGATGTG TTCTTTGAAG
```

25

```
      TGTTACCTTT  GTGTCCTTAG  AGGCTATGGT  GGGCCCTCAG  GCCATATATG  TTGGATGATT

      TAGTGGAAAA  GAGATTAGAC  TTTTATATTT  TATTCTATTT  ATTTATTTAT  GTTTTAATCT

      TGTGTAAGCC  TATAGAGGCC  AAGATAGGGC  TCAGAGTTAA  AAACAGTGAT  GTTCTAACTT

      TCAAAACTAC  TAAAAACAAA  ATTAGCTATC  AAGGGAAATC  CTGAGTGTTC  TGACTGCAGG

  5   GGTTCCCCAA  AGCCTGGAGG  ACCGAAGCTT  TGAAATGCTG  CATGGTGTGC  CTCACCTCTG

      TTGAGTATAG  TGAGGAACAT  GAGCTGAGGA  GTGGATTAGA  TGAACTCTTA  AGATTTTTTT

      CCAACTTTAC  GGTTTGAGTT  TATGAGTAAG  TAAGTGAGGA  ATGAGCCCAG  GGACTTTGAC

      TACTTTGCTG  TGCAGTAAAT  TGCTGTTTTC  TTAGCTAATT  CTGGAGTGTG  TATTTTGGTT

      TTAATGTAAT  GTCCTTTCAG  CACTGGAAAT  AAACAGATTT  ATGACCCTCC  ATTTTCACAT

 10   GTAACTTCCA  GTTCCTTCAG  TAAAATGCTG  GAAGTTGCTT  GGCTGTAACC  ACTGTAACTA

      TTTCCTGGCA  GTTTTGCATT  TTATTACAAT  TCGACTTTTT  AAATGCATGG  TTCAGTCCAG

      TTAATAGGCA  GAAAACTATA  GCTATTCCTT  GGAACTGACA  TAGGGAAGTT  GGTTTCTTTC

      TGAGGAAAGA  TGTGACACTC  TGATGGATTT  GGCAAAATAC  CAATTCAGTA  AAATTTGCCT

      GGTTCCTTTC  TGACAGCAAA  TGGAGGTAAT  TGTACGTATA  TAGATTTCTA  CAGTAAACAA

 15   AGGGCAAAAA  TGGAAACATT  TTTTGTTTTT  TAATGTTGCT  GGAATATAAT  ATTTCCATAT

      TTGGTTAATC  ATCAAATCAT  TCAACTCCAT  GCAAAGTACT  CTTTTCAAAA  TTACAAATCA

      TCTTGATGGA  TTAGACTTGG  ATGTCTAAAT  TTTTTATTAT  GTATTCTAAA  TTATCTATAG

      GTACATTTTG  AAAATTATAT  ATACATGTAT  CACTGTACTA  CATTTGTTAT  AAAATGTGAA

      ATATTAAGCG  ATTAAATACT  TTATTAATGT  CATAAAATTG  TGCCCTACCC  AAAATATATT

 20   TTAAGAAGAT  ATCATTTGAA  AATAATGTAA  AATATTCTTC  ATTATTATGA  TTTTGTTGCC

      AAAAGTAATG  TTAGAGTCCC  TGTTACTGAT  AATTTTCATG  ACTGTAAATT  TTCAACATTT

      GGAAACAAAA  TTTTCTGATG  TGTTTTAAAA  TCTTAAAGAA  TTTCAGTAGA  TTTTGACACA

      TTTTTGTTAA  AAGTATGAAA  GTGTCACAGC  TTTAGAAGGA  TTTAGAATAA  TTATCAATTG

      GTATTAGGGA  TGACAATTTT  AATAGACAAA  TTTTAACTGA  AACATTGCT  TTATTGATGT

 25   CTAGGACTAA  AAAAAATATT  ATAGTGATGG  AACCATTTCA  AAACATCTGT  TTTTAAAAAC

      TTCTGTCAAC  ATGTTTGACA  GAAGTCAGAA  TAGGGTAATT  ATAAAAAAAC  TGCCATTTTC

      TTATGTTTGC  ATCATTAGTA  TTTGCAAATC  CTTGTTTCTT  TGCAAGTATA  TTTTAAACTA

      CTTGTTCATT  TTATGAGTGT  TAGCCTGATT  AATTCAGATA  AGATGCTCAA  GTCAGTTTAA

      GCAGGCAGAT  GTGAAAAGG  GAATATGCTG  TGCCTTAACT  ATCTTATAGG  GACTTACTAG

 30   CACACTGTAC  ATAACAGATG  AACATCTGAT  ATTAGATCTG  ATGGGCACAC  CAATGTTGTT

      TCATATATTA  AATAATAGTG  GAAGTTTTCA  ATTTAATTAG  TTTATTTATT  TATTTATTTA

      TTTTTGAGCC  GGTGTTTCCC  TCTTTTTGCC  AAGGCTGGGG  TGCAATGGCA  CATGGTTGCA

      GCTCACTGCA  ATCTCCACCT  CCCAGGTTCA  AAGGATTCTT  CTGCCTCATC  CTCCCAAGTA

      GCTGGGATTA  CAGGCGTGTG  CCACTATGCC  AGGCTAATTT  TTGTATTATT  AGTAGAGACA

 35   AGGTTTCACC  ATGTTGGTGA  GGCTGCTCTC  AAAATCCCGA  CTTCAGGTGA  TCCACCCGCC

      TCAGCCTCCT  AAAGTGTTGG  GATTACAGGC  ATGAGCCACT  GCGCCTGGCC  AATTGTATTT

      TTATGTTAGA  AATATATATG  GGAGTTACAA  TACTTTGTTT  CCAATGTTCC  GCTGGAAACC

      ACTGGTTGAA  AGGATAGGTC  CCAGTGTCAG  CTGCCCAGTT  GGCAGTTTTG  GATAGAATAT

      AAGTTCTATG  AAGGCATGGA  TGAAGTATCC  TTCATGGTAA  TGATGGTAAT  GTTAGTATAG

 40   TCACTTGTTT  AATGTTGTAT  AACTTTTTAA  CTCTTCCAAC  TTCAATTTCA  TCTATATCTT

      TCATAGCAGT  CCAGTAAAGT  AGTATCAATT  ATAGATATCA  TTCCTACTTT  CAGATAAAGT

      AACTGAGGCT  CACTAGTTTT  GACTTGCTCA  AGGTTACGTA  AATGATAAGT  GACAGACTGG

      ATTTAAAGGG  CGGAGTCCAC  TTTACTAAAC  TGCCTCTGTC  TTCTAGCCAA  GTTTGTTATG

      TAAAGTAAGT  ATTCAGACAT  TGGCTAAGTA  TATTATTAAT  CAAGAAAACA  GTAGAACAGG
```

```
      TTTAAGGAAT TTTTTTTTCC CTCTCCTCAG TGTCAAAAAG CAGTCAAGTA AAAGATCCAA
      GATAAATAAC TTCAGTTCCC AAAAATAACT TCAATTCCAA AATTTCCCTA TGACTAGGTA
      TATCTAATTG GATAGCTGAT AAAATGAGTG GGAAATTGAA GAAGGGATAT TTTGAGAATG
      TGTGGCAACT TAATGGTAAT ATTGAAGAAG CAAATGGATA GTTATTGAAT CATAATAGTT
5     TTTTAAACTT AAATACTAGG GCTTGATCTT CATATTTTAG GAAATTATCT ATTTGTTGCA
      GTAGAAATTA GCATAAATGA ATGGAAACTT AAATTATCTC ATCTGTCAA GGCAGTCTTC
      AGAGACTGAT GTTTATATAA GTGAAAACTC AGAAAAATTT TTTGGCTGCA TTTCAAGCTG
      GATAGCGGGG TCAAAACTTC AACTTCTTAA TTGGTGATAC AACGTCAGAA ATCACAAAAG
      AAGTATTTCA CAGGAAAAAA TGACATACCC CGAAACATT CAAATCAAGA AAATCACAGG
10    GGTACAGTGG ATGGTCCCCC TTAAGGTACT GCGTACTTGG CTTCCCCGCT GGTAAATGCC
      TCTCTTTTAC TAAGTTGTTG TCGTGCAGAG TGGCATTGCA GCTCTGTACT TTTACAGAAT
      AATTTCAGGT TCTAAAATTG ATACATTATT TCAGATTGTA AGAATAGTAT AAAATAAAAT
      TTTAAAAGGG AAAATAATTC CATCAATTTA ATGGAATGTG TAGGGTTTAA GTTATAACAA
      CAAAAACAAA GTTGTAGTTT TTTAGGATTA CTCATATAAA TAGTGTGTCT ATTAAGAATT
15    ACTGGCTTTA TGAATATTAA ATAAGAAGGC TGGCCGTGGT GGATCATGCC TGTAATCCCA
      GCACTTTGGG AGGCCGAGGC AGACGGATCA CCTGAGGTCA GGAGTTCAAG ACCAGCCTGG
      CCAACATGGC GAAACCCTGT CTCTACTAAA AATACAAAAA TTAGCCGGGC GTGGTGACGC
      ATGCCTGTAA TCCCAGCTAC TCAGGCGGCT GAGGCAGGAG AATTGCTTGA ACCCGGAAGA
      CGGAGGTTGC AGTGAGCTGA GATCGCGCCA CTGCACTCCA GCCTGGGTGA CAGAACAAGA
20    CTCCATCTTG AAAAAAAAAA AAAATTAAGA CAACCACAGA TATAAGGATT TGCAATTAGA
      AAGCTTGAAG GGAAATTTTT GGAACTTCTA GGTCAACCTA AAATTACAAA TAGGGAAGCA
      GCTTTGGAGA GGAACATTTG CACTTCTGAA GTAATACTGT GATTTGGACT GTCAGACATT
      GTGATATATT GGCATCACCA CACCTCTGTT AATTATAAAC TTACCTATTC TCAGTGCCTT
      GTTAACTGTT GAAATAGGAC TTGACATTTG TTGTTCTTCC TCTTCTGTTA TTCATTTATT
25    CACTCATTTA GTAGTTAATT ACTTCCTAGT GCCAGATAGT TTTCTAGGTA CTGAATAAAC
      AAAAATCCTA TTGTCATAGT TATGTTTTTG TGGAAAGAAC AACCTGATAG ACTATCGTGA
      AAAATAAATG AGTTAGTATA CATAAACAC AGCCCATGAG ATATAGTCTG TAATTATCAT
      CCCTGCTTCA TTTATTTATT TATTTATTTA TTTATTTAGA GAATGGATCT GATTCTGTCA
      TCCAGGCCGG AGTTCAGTGG CTGGATCATT GCTCACTGTA ACCTCAAAAC ACCTGGCCTC
30    AAGTGATTCC CCCAACCTCG GCGTTCCAAA GTGCTGGGAT TAAAGGCATG AGCCACTGTA
      CCTGGCCATT GCTCCTGTTT TAAAGATGAG GAAACTGAAA GTACAGAGAG GAACGTGACT
      TGCTCAGGAT CACACAGCCA ATCAGTGGCA GAGCAGTCTA GGCAGTCTAG GCCTGAAGGC
      ATTATTCTTT CTTTCTGTTC TGCGTCAAAA ACCCTAGGCA GGCATGGTAA AAAGACTGAA
      AGAGCCAAGG CAACATAGGT ACCTGAAGAA TGGGAGAACA CCTATGAGGA CTGCTAGAAA
35    TTTTAGGCAG CCCTTTAAAG GCCTTGGGCA ACAATTTGAA CGTGACTCAT GAGTCAGTTA
      ATTCTCTGCA CCTTTTTTTT TTTAGAGAAT ACAATGGATA AAGCAATGAC CTTTGTAACA
      TGTAAACCTG TATTTGAATC CTTTCCTCTC CATTTAATTT TGCAGAGTCA CCTTTTCTTC
      ATTAGAAAGA AATTTTATCC CAAAGAGATT ATCGTGAGGC TCAAATAATA TAGGTGGAAG
      CCCTTTGGGT TTAAAAAAAA GCCTTTTCTG TCATTTCTGG GTATCTCCCG TCAGTGATTT
40    TAGTAATTGT TTACCTCTCT GGCTCTTCGT GAGAAGACAA ATTGTAGGAA GAAGTGGAGG
      ATGGCATTGG GTAGGATGCT TAGTTCTAGT TCCACCACTT CCTTAGGTCT TTTTTTTGTC
      CCCCAAGACT ATATCATTTC TGTGTTCTCT TGAACCTGTA AGATTTATGA GTAATCTGTA
      CAATATTTAT AAAGTATGAA ATCTCTGCCA TTGTGGCACA AAAGTAGCCA TAGACAATGT
      ATAAATGAAT GGGCATGCTG TTCTCCAGTA AAACTATATT TATAGAAATA GGCAGTCTTC
```

27

TCCAATTTTA GTTTGCTCAC CCCTTTCCTA TAGAACTCAG GTTTATCATA TACAATGTTA
CACATTATTC CTTTGAATAA ATTGTTTACC TAATACTTTT TAAATTTGGG CATTTTCTAG
GACATTCTAT TGTTTTTATT TTTAACTTGT GTTATACTGC AGTGAATGTC TTCCTGTGTG
TAGCATTTTG TTTGTGTATG TGTGTCAATA TCTATAAAGT AGCTCTTAAG AACTCTGGAA
5   ACATTGTCAG ACTGCTTCCT CAAACAGCAG AGGTGGGAGG GAGCATCGGC CTAACCCTAC
CTGGCCTGCC TTACCAGCAT GCAACCAAGG AGAAGAGAAT AGGATCAAGA GCCAGTGCCA
GAGAGAAGTA CATAATACAA CTTTATGGAG TTTAGGTTGC ATCCTATAAA GTTACCATGT
TACACTGTTT CGATCTACCA AAGTAGAAGT TTTACATAGT TTAGTATGAA GTTAATTTTA
TTAACCTTTT CACACAACAC ACCCTCTCAT ATTGCTGCCC AACAATGTCC TGGGTCTTTG
10   TTTCCATGGT CATCATATTG AGTAGGGTTT TAGCAGGGAC TTGAGTAGTT GGGAACTGAA
CTCATGCTAG TGAAGCCATT CTTCACTAGA AGCAGACATA TGGGATCCAG AATATCATGA
ACAACCAGCT GGGTACAGTG GCTCACACCT GTAATCCCAG CACTTTGGGA GGCTGAAGGA
AGCAGATCAC CTCAGGTCAG GAGTTTGAGA CCAGCCTGGC CAACATGGTG AAACCCTATC
TCTACTAAAA ATACAAAAAT TAGCCAGGCC TGGTAGCACA TGCCTGTAGT CGCAGCTACT
15   TGGGAGGCTG AGGCAGGAGA ATCGCTTGAA CCTGGGAGGC GGAGGTTGCA GTGAGCCGAC
ATCTCGCCAG TGTACTCCAG CCTGGGCAAC AGAGTGAGAC TCTGTCTGAA AAAAAGAGAA
TATCATGAAT ATCACCAAAG ATAAAGCAA GATGCGGATA TGAGTAGCCC AGAGTAACTT
TTCCTGTGCT ATATAACACA GTGAGTGGTG ATTCCCACAA TTCAAGTCCA GTTTTTCTGG
ATTCTAAAGC CATACATTTT TAACTCTCCC CCTAAATATG ATTTGAATTC AGAGAGTCAG
20   ATGCAATAAG GTAATTTTAA CTTTAAGTGT AGTTTAGCCC ACACATTTCT CTAAACTGGG
GGTTGGCAAA CTACAAACCT CTTTTTGTAA GTAAAGTTTT ATTGGAACAC AGCCTAGCTC
ATTCATTTAT GTATTCCCTT TGCACCTTTT AGGTTGAGTA GTTGTGACTG AGGTCTTCCA
GCCTTCAAGC CTAAAATATT TACTATCTAG TCCTTCAAAA AGAAGTTTGC CAACCTCTGC
AAACAAGTGA AATAATGTGT ATTAGAGTAG AGCAGGAGTC CACAAACTAC AACCTCTGGA
25   GCCAAAACCA CCCTACTGTC TGTTTTTGTA AACAGAAATT TTATAGAAAC ATAGTTGACC
CCTGTGCCAA TTTGTTTGAG TCTTGCTTTT GATGGCTTTT GTGCTACAAG GGCAGAGTTG
AACAGTTGCA AGAGAACTAA GTGGCCTGCA AAGCCTAGTG GTCTTTGCTG TCTGGCCATT
ATAGGAAGTG TTTACCAAAC CCTGCATTTG ATCATGGAAT CCCAGAATGT TTATACCTTT
GAGGATTTTT TTTTCTACTT TCCATATTTT ACAGGCAACA GAACTAAAGC CTCGTAAAAT
30   CTTGCCCAAG GTTCTACAAA TAGTCTTTTC TATTAAATTA AGTCAGAGCC TGAAATTGTC
TGTGCCCAAG TCTCTCCTCA GAAATTTAGC ACAGCATTTT CCCTTTCTGT CTCCCACATG
CTACACAGTC ATCCTGGATC ATGTGCCCCG CCTCCATGAG GAGATGTTCT AACATGAGGG
TGACTAGGCT CTGGAGCCAG AGCATCTTAG TTCAAATCCT GGATCTGTTA AAAATAATAG
TTATTTTTCA GTTCTAACAC TGAAATAAGA CAAAAGGCCA TATTACTGAA ATAAAGTGTA
35   AAGAATATAT TTAATACTCA GGTTACTTTA GCCCCCTCAT GATCAGAATA TGCTAATAAA
TTGAGAGATT GACAGAAAAA AAGGATTCTG TTTCACTGTG GCTCACACCT GTAATCCCAT
CACTTTGGGA GGCCGAGGTG GGTGGATCAC TTGTCTGGAG TTCAAGACCA GCCTGGCCAA
CATGGCAAAA CCCCGTCTCT ACTAAACAAT ACAAAATTA GCCAGGCGTG GTGGTAAGCA
CCTTTAATCC CAGCTACTCG GGAGGTTGAG GCAAGAGAAT CGCTTGAACC TGGGAGGTGC
40   AAGTTGCAGT GAGCCAAGAT CACACCACTG CACTCCAGCC TGGGCAACAC AGCAAGACTC
CATCCCTGGA AAAAAAAGT TCATCATTTG TTCTTGTAGT TTCCTAGTGT GGCCCTTGTC
TCAGTGGAAA ATTGTAGTGT GATAGACTGA GGAGAATTCT GTTTCTGCCA CTTCGCCTGT
TGCCTTGGTA GGAGCTCCTT ATCTCTGAGC TGAGCTCCTT TTGTCCTTGG TAAAATGTGG
TTATTATCAT CTGCATCCTC ATAGTCTTCT TGTAAGGACT GTGAAGTCAC TTTTGAAAAA

28

```
TGCCTTGAAT ATGCCAGATG GTCTGTTTTC TTTCTTTTTT GTTTTGTTTT GTTTTGTTTT
CCCTTGGCCT TCATTCCTGG AAGTGTTTTG TTATATACAT TTTGTTTGGT GGCCAACAGT
GCTAACCACA GAAAGTTTG TATGTCTTCT TTTTCTAGAA TAATTCTTAC AGTGTAACCT
CTCAGGGTGA AGTTTTTGTT TTGAAAGGGG CACATTTACT GTGAATGAGA GCCATGGAAA
TGGACCTGAA ATTTGAAGAC AGATTTTACT TGGTTTTCTT TTCTCTTATT TGTTTGCAGT
ATAAAGAAGA GGGCAGCAAG GTGACCACTT ACTGCAACGA GACAATGACT GGGTGGGTGC
ATGATGTGTT GGGCCGGAAC TGGGCTTGTT TCACCGGAAA GAAGGTGGGA ACTGCCTCTG
AGAATGTGTA TGTCAACACA GCACACCTTA AGAATTCTCA GGAAAAGTGA GTTGCTACAA
ATGAGACATA CGTTTAGTTT TGTTTTATAG TATATATGAA TATGTGTGTA CATTTTTGGA
ATTTTAGTTT GATTATACAA AATATCTTTG GCTTAGAAAT ATTAGGCATG CTATGTAAAA
CCTTACTGGA AAAATAAATT GACCAACATT ATTGAGAGTA TTTTTTCAAA GTGTTCCAAA
AGTAATGGAC CAATGATTAC TTTAAATGAA ATCATGTAAT GGACCACAGA ATTGCAAATT
ACTAATAAAG AAAAGCCATT TTGCTTATTG CCATGTAATA ACATGTTGCA TGATACAAGT
AGATACGTAT GTTTATGCTG CAACAAGTAT AGGTGATACT AATTGGGCAA CTTTTAAACA
AGACCATAAA TAACTGAAAT CAAAGTTCTT AGTATTTATG CAGCCTGTTG GTTTGCGAGG
GCTGCCATAA CAAAGTACCA CAGACTGGGT GGCTTAAGGA AGGTTGTTTT TTCACAGTTT
TGGAGGCTAG AAGTCCAAGA TTGAGGTGTT GTCAGGTTTG GTTTCTTCCA AGGCCACACT
CCTTGGCTTG CAGATGTCTG CTTGCTTACT GGGCTCCTAA ACAGTCTTTT CTGATTGTCC
TAATGTCCTC TTCTTATAAG GGCATCAGTC ATATTGGATT GGGGCCCACC CATATGACCT
AATTTTACCT TAATGACCTC TTTAAAGCTC TGTCTCCAAA TACAGTCATA TGGCTGTGAT
ACAGGGGGTT AGAATTTCAA CGTATGAATT TTGAGGGGGA CACAATTCAG ACTATAAACT
GCAGTTAATG TTTACTGTTA AATTAAATCT ATAACTGACT AAACCACACA AACAGGAGAT
TTCAAATTAG ACTTTATTAG TTTTGGAAGG AAGAGATGAA ATGTTGTTAC TTTTTCTGTT
TTATGTAGTT GTAAAAGCCA CTAAAAATGT ACATGTACAA ATCATCCCAA GCCAGGCAAC
ATTAGTAATG AATGACTGCA GCACAGAGGT AAGAGAGATG ATTTAAAGGA GGATGAATTG
TCTGCAAAGG TGCTGGGCAA GATCTAGACT AATTCATCCC CTTCATTTTA AGTTTAGGTT
TTAAATAGCT TTGTTTGGCT TGATCCTAGA GCTACACATT TACTTTTAAC TTGTTTACTT
TGTACCCTAT CATTTAGGAT ATGCTATGTA CTATTGTACT CTATTTGATA TTTCAAACAT
TCTCTCATTT AGTGAAGAAG CTCCCAACCG GAGTGCTCAG AACCTACATT GCCTCACCTG
TGAAGTGAGC ATGTGGGACT GAATTGCCTT TGAGGTCTCT TTCAGCTCTT AATGATCTGT
TATCCCATAG TAAGATACAT TATTTTTAAT CTCGTTGGAT CTTAAACACC AAAAATAATA
GTATTTAAGA CGTAGGATGC TATCTTGTCA TTATTTTAAT GCACATGTCA ACACACGAGG
TTTTGCTAGA TGTTATGATA AAGCAAGCGA AACAGCAGGC TACTGCTCCT CAAATATCTA
CAGTCAATGA AATATTGACT GGCGTTGTGG GAAAATATCT CAAAGTATT TTATTTACAA
TTTAAAATAT ATTCTTTTGC CTAAAGAAAC AATGAAACAA CTATAACTTG TTTTTGTTAT
TGTGTTAGTG TAGTCATACA TATGATACAG TTTTATAATC TGCTTGATAA GGGTACGTCT
CTGGGTTTTT CACACTTTGC AACAAGTGGG AAAGAACTTC GACAAAAACA TTAAAACATT
ATCAACCTGG TTGCACGTTT TTCTGATTAT TTTCAGTGCT CTTAGGTTTT TAGACCATAT
CAAAATTCTC TCTCTACACA CATGATTTAT AAGGGAAAGA AAAGTTTGCA GTATGCCTAA
AATGTTCCTC AGGTTAATTA TATCTTGTTT TGCATAGTTG AATGTTTATG GAAGTCTTAG
ATAGTGAGTC ACTCTGAAAC CACAGACTCT TTTTTAATAG TAATAATATG GTGAGTTCTA
TAATTCCTAG CATCTAGCAC ACAGGTATGA TAAATGTTTT AATTAAGTGA ATAATCAGCT
TCCTGAATTT TTCTACTTTG TAAATATGCT CTTTGTGTGT TGAATCTATT TACATATATA
TCCAAGCATT TGGCACAAGG TAAACCAATG TGGCTCAGAT TTTAATGTTA TTGTGAATCA
```

29

```
       CTTGGGGATC TTGTTTAAAA AAAAAAAAAA AAAGGAGATT CTGATTCAGG TGATCTGGGC
       TGAGTACTTA CAGTATGCAT TTCTAATAAA CTTCTGGGTG TTGCTACTGA TATTGGTTCA
       AGAACTAAAT TCTGAGTAAC AAGTATAAAC CGAATGTCCA TAAAGTATAA TTTCTTTTAG
       TATCATATGG TTATAATGCT TGCATTTTGT ATATATGTTG CATACAGGTT TTGCTTTTCT
  5    CTTTTTTCTT ATTAAAAAGG TTATAAAGGT ACACTGTAGA AACTGTAGGG AGTAGGGGAA
       AGCTTGGCAT CTCATGCCAC CATTCAGAGG TAACCATGTT TGGCATTTTA AAAATGCATT
       TATAGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTCTGC GCGCATACAT ATATATGTAT
       GTATTATACA TATACGTTAA TATACATATA ATATACATAT ATTACTGTGC AAAAGTTGAT
       AAATTGATTA TGATTATGTA TAAACTTTGA TATCTTCTCA TATTTAATAT AATATAGTTG
  10   TGAGCACTCT TCCATATCAT TAAATAACTT TCTAGACATT TTAAATGAAT ATATAGAAAA
       TGAAACATCA TTAATCAGTC CCTTATCTTA GAATATTTAT TTCGGTTATT TACAATCTAT
       TAGTAGTTTA AAAGTACTTC ATGGAAAATC CTTATACATG CATCTTAGTT AACTTGAACT
       TTTATTTTTA GAGAAGACAT GCCAGGCTGT TCTAGGCTAT TGATACATAT TGCCAATTAC
       CTTCCTGAAA GATTGTACCA CTTTCCACTT AGGCACAGAG TGTGAATGCC TGTTTTAATT
  15   ACTCTATTTA CATCAGTAAA GACTATATCT GAAATTATCT CATTTTTGGG GACAGGTATT
       CTAATAGGCT CTACAAGTAT GATCACAACT TTGTGAAAGC TATCAATGCC ATTCAGAAGT
       CTTGGACTGC AACTACATAC ATGGAATATG AGACTCTTAC CCTGGGAGAT ATGATTAGGA
       GAAGTGGTGG CCACAGTCGA AAAATCCCAA GGTAATCAAG CACACATTTT ATCATTAATA
       AAAATATGAA TGCTGAATAC CATCCTCCCT CCTAAGCAGT CCTACCAGTG TTGCTCGCCA
  20   TTTTATTGGT CATCCCAGTT GTAACTTTTT ATGTGATCTG TTAACGATCT TTATCTCCTA
       TTGATACATA TTCTGTCACT CTCCAAGTCC TATTTATGAG TTTTTCTTTC ACATGTCTGT
       TGTTTCCATT GCCCCCTTTC CATTTTCTAC ACTCTGGCAT GCAGCTTTGT CCTGTTTCCC
       CCATGCTATG GATTAGCCTC CTGCCGGTAA CCACCTGAGG GGTTCTTCCT GCCTGCTGCA
       TAAAGAAAAA CCATGGCAAT ATAGTAGAGA AAGAGCTTAA GAGACACTAG GCTGGCCATG
  25   CCACGTGGGG GATGGAGTTC ATATTCAAAT TATCTTGTCC AAAGCTCATA GGTAGGGGTT
       TTTCAAAGGC AGCTTGGGGG AAGGGGTGGG GGTGGCCAGG TAACAGATGC TTGCTGCTGA
       TTGGTTGGGG TGGGTGAAAT CACAGGGAGT TGAAGCTGTC CTCCTGTGGG CTGAATTGCT
       TCTAGGTGGG GCCATAGGAG TGGGGTTGCT GGGTCCAGGT AGAACCACGG GTGTCAGACA
       TGCAAAAATA AAATAAAATA AAATAAGATA ANGGTAAGAT AAAATAAAAT AAAATAAAAT
  30   AATAAATAAA ATAAATATAA AATTTCCCTG AAAAGATATT TCAAAAAGCC AGTCTTAGAT
       TCTACAATAA TGATGTTATT TGCTGGAGTA ATTGATGGAG TTGCATGTCT TATAACCTCT
       GGAATAACGG CTGACAATCT CTCAAGTCTG CGCCTTAGCT GGACTCAGGT TCCTCTTCTC
       CCCACAGCCT GACTGCCTCC ATTAGCTTCA CAAAAGTGGT TGGGTTTCAG GGCAAGGCCC
       ATTGTCATTT AAACTGTAGC CGAAATGACT TCCAAAGTTA GCTTGGCCCA ATAGCCCAGG
  35   AATATTTAAG TGGAAGGCAA GATGGGGGAT GGGTTAGCTT AGCTCTCTTT CACTCTCATA
       GTTTTCTCAC TGGTATAATT TTTGCAAAGG CGGTTTCATG CCTGCCATCT CTTTCGTCGC
       TACCTCTCCC AGTTCCCATT CTTAGCTGTT TTATGAAATG CTTCTAGTTT CATCCTCTTA
       TACCAAGTTC TGGGAGACTG ATTTGAGTAA TAATAAAACT CCAGTTCCC ATACAGCCGG
       CTCTGCGTGA ATTAAACTCA TTTTCTATTG CAATTTCCCT GTCTTGATAA TCAGTTCTGT
  40   GTAGGCCGTG AGGAAGGAGA ACCCGTTGGG TGATTACGAG ACTGTGTTAC TGCCCACTAC
       CTAATGGATA CATTTAGCCT GGTATCCAAA CCCATCTAAT TATGACCATA ACTATATTTA
       TCACCTTGCT CTGCTTTCAA ACGATATGAC ACACAATGAA TGAAACTTT CATTTTTCAT
       CTTCATTTGT GCTGTTCCCT TTGCCTCAAA TAGCCCTCAA CTTGCCTACA GTAACTGTAA
       AATTTGCCAC CTAAAAAAAA ATCTCAAAAT CCTCTCTATG CTTTGATGTC CAGCAAAAAA
```

                                      30

```
AAAAAAAAAA AAAAAAAAGT TATCAGTATA GCTCTATTAC TCTCAACCAG AGGTGGCATT
TTTCATCTTT ACCCATAAGC CCCATGGTAT ATTTCTTAGT TTGGGAAATT ATATCATAAT
ACTTTGAATC TGTCTAGTCA GAAATTTAGA TAATTTAAAA AAAAATTATT TTTCTGAAAG
TATAGCTGGA TCTAGCTTAG GATCTACATC TATCTAATAT AGTTCCTTAA CATNTTGTTA
AATGGCCACN GGNATAAGTC CTGTAATGCC ATACTTTGCT TTAGGATCAT GTGACTAAGG
GGTAAGGAAT TGGAAAGCAA TGGGGAGCTA GCAGAATTTG AAAAATTATA TAGGTAGGTT
ATTTTTCCTT AATACATTGA AATAGCCTCA AATTCTCAGA GAATACAATG TTTAATCCTC
TAATATCTGT AGAGTTCATG GCTCAAATTT GTATTATTTG AATAAACTAC TAATAGATTA
ATAATTACTC AACTAAAACA CTTTGAAATG TGAGAGTTCC CTTATCTCCC TCGCAGGCAT
ATGACAGAGG TGTGGCTTCT CACACCTTTG GTTGCCCTTG CCCTACCACC CAAACCCCTA
GGGGGAGCAT GCAGAGGGGC AGGTGCAGAG GCCATGGGGG GTGCTTTTGG GCTCTGGCCC
CACAGCAGTG TCTAGGAGTG GATGTTGGAG ACTCCTGAAG CCCAAGTGGG CATGTGTTAC
AGTGTGCTCT TTCAGCTTAG CCGTCTGCAG ATGGCTTGTG TTAATCAGGT CATTAGACCC
CATGCCTTAT TGCAAGGGCA GGGGTCCAAT GTGACAGCCT AAGTTCTTGC TCAGTGTACC
AGAAGAATTG GATCACACGT GGGCTGGAAG GATGAGCACA AGGTTTTATT GAGTGGTGGA
GGTGGCTCTC CGCGAGACGA CTCTCAGCCA GAGAAGGAGG ATGGAGGGGA AAGGTGTTCT
TCCCCTGGAG TCCACTTGTT CCTGTTCTCC TCTGGGTTCA AATACCTCTT CCCTCCTTTT
CTGCTGTGCT GTCCCACCAC TCTCCACAGC TCTGTGCCGC TCTGTTCCTC TGCTCCTCTG
GATGTTCAGC TCCTTGTATC TGTGCCTGCT AAGATTTTGG GTTTATATGG GGGAAGGATG
GGGGGCATGG CGGGCCAAAA GGCACCTTTT TTGGTGTGAA AACAGAAATG CCTGTCTTCT
CTTAGGGCCT TAGGTCTTCA GGCTTGAGGG TGGGGCCTTT GCTGAGGGAC CACCCTCTTC
TACCCAGTAT TTCCCTGTCT CCTGTCCATA TCAACATTAC TAACTTTTTC ATCTGCAGAC
TAATAATGCT AAGGTGTGGC ATTTTTCAAC TGTGAGACTA TGTGAAGGTT TTTCCTGTCC
AACTGATGGC ATCCTCCCAT AATTCTACCC CTTTCTTAAA AGAATCTTTT GCAGTATTTC
TCCAAGTTTA TTCTAGAGAA TTTCTTGTCT GTGAAATGCT CTAGTTAATC AAAATTGGAA
AACGGAGCAT ATCATATCCC CTTCTCAAAT TCACCAAAGT GAAGTCCTAA TGTGTCTTAA
TGTATCTGCA TGAGACAGGA AGCTGAGATC TATTCAACAA CAAAAATCCA AACAAGCATC
AAGAGGAGGA GTGTTAGCAC TTGAGCCTAG GGAGACTGTG GCTCCTGCCT GAAAGATGGG
AGCCTCAGTC ACAGCTGCTT TACCAAGTGT CATATGCTAT GTTTCTGAGG ACTCCTGCTA
AAGCTCCCTT CTCCCTCCAG CCAACCACTT TTGTTTTAGA CAAGGGCTGG GTTTATGAAG
GACTGTTTTC ATGACTAAAG CTTTATAGAA GGTTTAAGAT AAGGAGATGG AATTGAGTGA
AGTAGGAAAT ATGAAAGCAG ATATTATAAT CTGGCTTCCT GATTTTTCAC TAGCATTTTT
GTTTATAAAT TAGTTCTGTT CTAAGAATCC AATGACGTAA TAGAAACTCT CAAAGATTCT
TAACTTGAGA TATAGGGAGT CTTTGAAACT GCTGAAATTA CAGACAGCAT TTATTGTTTA
TGAGCATTTC TGAATCTAGA GCTTTCACTA GATTGTAAA GAATGTGGGC CAAAAGATTA
AGAGCCAATC CTGTATCTTG TACTCAAAAT GTTTGTAATT CTCACCTTTT TATCCCCCAG
ACTCCTCTTG CTCTCTCTTC ATTTTCACAG TGTTACTGGA AAGGGGTCCT GATCCAGACC
CCAAGAGAGG GCTCTTGGAT CTTGTGCAAG AAAGAATTTA GGGCGAGTCC TCAATGCAAA
GTGAAAGCAA GTTTATTAAG AAAGTAAAGG AATAAAAGAA TGGCTACTCC ATAGACAGAG
CAGCCCTGAG GGCAGCTGGT TGTCCATTTT TATGGTTATT CTTGATTGT ATGCTAAACA
AGGGATGGAT TATTCATGCC TCCCCTTTTT AGACCATAGA GGGTAACTTC CTGACATTGT
CATGGCATTC TTTTTTTTTT TTGACAGAGT CTCACTGTGT GTCCAGGCT GGAGTGCGGT
AAAGCAATCT TGGCTCACTG CAACCTCTGC CTCCTGGGTT TAAATGATTC TCCTGCCTCA
GCCTCCCAAG TAGCTGGGAT TACAGGCATG ACCCACCATG CCCAGCCAAT TTTTGTATTT
```

```
        TTAGTTGAAA CAGGGTTTTG CCATGTTGGC CAGGCTGGTC TTGAACTCCT GACTTCAGGT
        GATCTGCCTA CCTTGGCCTC CCAAAGTGCT GCAATTACAG ATGTGAGCCA CCAAACCTGG
        CCTGTCATGG CATTCTTAAA CTGTCATGGT GCTGGTGGGA GTGTAGCAGT GAGGAAGACC
        AGAGGTCACT CTCATCGCCA TCTTGGTTTT GGTGGGTTTT AGCTGGCTTC TTTACTGCAG
  5     CCTGTCTGTT CTATCAGCAA GGTCTTTATG ACCTGAATCT TGTGCTGACC TCCTATCTCA
        TTCTATGACT TAGAATGCCT TAACTGTCTG GGAATGCAGC CCAGTAGGTC TCAGCCTCAT
        CTTACCCAGC TCCTATTCAA GATGGAGTTG CTCTGGGTCA AACACCTTTG ACAACATCAT
        TAAGCCTCAG TTCTCACACT GTTTTTGTTT TGTTTGTGTG TGCAATGATG GGCAAATCTC
        TGCCTTATAG GATGGTAGAA AAAAGGAACT TAATATTGTA ATGACTGTTT TGTGCCAGAT
  10    AATTGCTTTA AACTGTATCA GCATCTTATT TAGTCCTGTT AATGATATGA ATGTTATCTT
        CATAACAGCT GCCATTTTAT TAAGGACTTA TCAGAGAAAA ACACTGTTCT AAGCACTTGT
        TACCCATTAT TACATTGAAT TTTCATAACA ACCCTTTGAG GTAAGCATGA TTATACCCAC
        TTTAATAGAA GTGAACTGTA GTTTTGTAAT GTTAGGTTCC TTGCCAAATG TTACACAGAT
        AGTAAGTGAT AAAATCATAT GCCCTGAAAT TACATTATGC TGCCAAACTT AAATTTCTTT
  15    TTTATCCTTT ATATTAGTAT ATTCTTAGGT TTAAACAAGA CAACTAGTTA ACACATACTA
        GATTTGTCC ACAGTTCCTG GCTCATAACT CCCATAGCCC TTGTCACTAT CTTTTAANCG
        TTGGGGCACA TTAGGCCTCA GAAGTAGGCC TCANGAAAAC AGAGTCTCTC TCTCTCTCTC
        TCTCTGATCT TCTTCCGCCC TCCTCTCACC TGCCCAGGGC AGCACTTTAA TCTTCTCCTG
        CCTTTCTGAT CTTGGGTCAT AAGACCTTCA TTTCCAAAGA TGTCCTGTGT CATACCCTAA
  20    AGGAAGGAAC ACTGAACAGA GAGAGGCTCA GAAGAATCTG GACAGGCCTT GCTGTGTTTA
        CATCATTCCC TTTATGTCCA GTCACATCTC TACATGGTTG TCAGTTGTGC CTATTTGATG
        AAGTCCCCAT ATAAGGCTCA CAAGGACAGG GTGCAGAGAG CTTCCAGATA GCTGAACAAG
        TGGAAGTTCC TGGAGGGTGG CGTGTTCAGG GAGGGCATGG AAGCTGTGTG CCCCTTCCCC
        CATACCTTGC CCTACTCATT TCTTCATCTG TTTCATTTGT AGTATCTTTT ATAATAAACC
  25    ACTAAACATT AGTTAGTATT TCTCTGAGTT CTGTGAGTCA CTCTAGCAAA TTAATTGAAC
        CCAAGGAGGG TGTCATAGGA TCCCCNACAT TATAGCTGGT TGGCCAGAAG CACAGGTAAA
        CAACCTAGGG CTTTCAATTG GCATGAGAAG TAGGGGGCAG TTTTGTGGGA CGGAGCCCTC
        AGCCTGTGAG ATCTGATGCC ATCTCTAAGT ACACAGTGTC AAAACTGGAT TGGAGGACAC
        CCAGCTAGTA TTCACTGTGA AATTGGTTGC TTGCTTGATT TGTGGGGAAA AACCCACATG
  30    CATTTGATCA CAGAAGTCTT TTGTGTTGAC AGTTGATAGT GTTCAGTGAG AGAATTAAAA
        AAAAATTGAG TTTCTTCTTC AACATACTCT CTCAATGTGA AACCACAGAA ACTATTTCCA
        TTCAAAGATG GAAATGGTTT GTTTGCATCT TAGTTTTTAT TTATACATCT TAGAAGAAAT
        GTCCAAGCTT TGTTTTTTCT CTCACCCTAT ATATAAAATT ACCTATGAGG CACAGATTTT
        TATGATCCTT GATTATATAG ACTTTGTCCA AATTGTGTGT TTTATAGCAT TACTGTAACT
  35    TGTTATAGTA ATCTTTGTGT ATATTATGTC TCTTAACATT GTCTTCCATA TTGTTAATGA
        CCATCTCATA TTTATCTCTG TATCATGTAT ATCTTCAACC AATGTGACTG GCTTAGGAGA
        AAAAATTAGT GAACAATTAA CTAGTGTTTG TGTAATCTAT ACAATTGTCA AGGTTACAAT
        TGCTATTTTT GAAGAAATCG TTGTTGTTTT TCTCTTTGTT TCATCTCAGT TCCATTTTGT
        CAAGGATTCC TTTTTTTTTT TTTTTTTTTT TTTTTGAGGC GGAGTCTTGC TCTGTCACCC
  40    GGGCTGGAGT GCAGTGGTGC AATCTCGGCT CGCTGCAAGC TCCACCTCCT GGGTTCATGC
        CATTCTCCTG CCTCAGCCTC CCGAGTAGCT GGGACTACAG GCACCCGCCA CCACGCCCAG
        CTAATTTTTT TTTGTATTTT TAGTAGGGAC GGGGTTTCAC TGTGTTAGCC AGGATGGTCT
        CAATCTCCTG ACCTCGTGAT CCGCCCGCCT CGGCCTCCCA ATGCTGGGAT TACAGGCGTG
        AGCCACCGCG CGTGGCCCCT TGTCATGTAT TCTTAACCTG TGTTATATCT AAGAGAAGGT
```

32

```
      GTGAGAGGCG GGACCATTTG TGGATAGGTG CAGAGGGCAT GAAGTAGCCC AGAAAGAATC
      TTTGCCATTG ACTAAATTTT AGCCTAGTAA AAAACATGTG GTAACCACTG AAAAACTAGA
      ACAGTGTGAT ATGACAATGC CCAGTTGAAT AAGAAATTCA AATAGTTTAT AGTAACAAAA
      AATAATTTTT ACCACATTGT GACTAGTGCC CTAAAAACAA ATTTGAAGGA GCAGAGAGAG
5     AGAGGAAGTG AGTTTTCGCT GGGGGCTGGT GAGTGGAGGC TACTTTATGA GCAGTTTTGA
      AAATTACAAG TTAGGGAAAA ACTCTTAGGG GAGTTTTAAG TGGTGAGCAA CAGGCACTTG
      AGGTTACAGG AGCAGCAGCA TAAAGATATA GAACAGAGAG GCCATTTGGC AGTCTTGGGG
      AAACTCCAAG AAGTCGAATG TGTCTAAAGC TGTGGTTCTC TCTCTGTGTG TGTGTGTGTG
      TGTGTGTGTG CGCGCGTGTG TTTGTGTGTG TGTGCATGTG TATGTATGTG TGAAATCGCT
10    AATACAGAGA AGCTGTTAGA ACAATTTTCT CACTTCCCAA AAATGTCCGT TCATTCATGG
      GCCTAGGCAA CTCTCCTTTG TGTGTCTTAC AGCCCAACAC TGTCATATAA GGTGTGGTTC
      TATAATGAAA CAGTTTTATC TGTTGTTTAC GGAGCTAGTC AGCCTGTGTT ATGCTTGTTA
      CTGGTTGAGG GTGTCCAGGT TCTTGGCTTC TTGAACAAAG AATTGGACAA AACTCACAAA
      TGAGGCAAGG AAAGAATGAA GCAACAAAAG CGGAGATTTA TTGAAAATGA AAGCACACTC
15    CACAGGATGG GAGTGGGCCT AAGTAAATGA TTCAAGGCCC TGGTTACAGA ATTTTCTGGG
      GTTTCAATAC CGTCTAGAGG TTTCCCATTG GTTACCTGGT GTATGCCTTA TGTAAATGAA
      GAGGGTGGAG TTAAGTTACA AAGTCATTTA CTCAGTATAG GCCTTGTGTT AATGGAGAGG
      GTGTTACTCC TGGGGGTTGT GGCCCATGTA AACGGAGAGG ATGAAGTGAA GTGACAAAGC
      CCTTCGCATT CCTGCCATTG CTGAAGTGTT TCCACTTTAT TTAGTTCTAG GAAGTCAGTG
20    TGAATTGGCC TTATGTTCCC TGCCTCCAGA ACCTGTTCTC CTGCCTCACA TAGACCCTTT
      TTTCCCTTTG CCAACTTCAC TTCTTTTACA GCCACCTCAC TACCAATGTG TCTGTCTCCT
      AAGTCAAACA TCAGTGATTC CTCTGTTTTC TCTAAACCCT TCTTATGTCT TCTACTTCTC
      ATCTCTTTCT TGGTCAAAAA TCTTTCAAAA ATGAGTAAGA ATGCAGCTAT TCAGGCAAAC
      TAAAAATAAC ATCACAGTGA TATACAAAC CAGTGTCATT TCACAAGGA AAATTATCAA
25    TACTAGATCC TGAAAAGAA ACAGCGAATG AAAGCCATTT ACACAACTCC ATTGTGTAAT
      TGACACATTG AATCACTCAT AAAACAGGTG CTCTGGGTCT GAATCTAGAT CCTAGCTAGT
      CTGGTAGCTG AAATCATAGA ATTATAGTAG AGTTTAGGAA ATCATCCTCA AAGGAAAGAT
      TATATGTTGA TATCAAATGT ATATTTCCTT TCTAGGCCCA AACCTGCACC ACTGACTGCT
      GAAATACAGC AAAAGATTTT GCATTTGCCA ACATCTTGGG ACTGGAGAAA TGTTCATGGT
30    ATCAATTTTG TCAGTCCTGT TCGAAACCAA GGTAAAAAAA TAAGCCTAAG TTTTTTGTTA
      ATTTGTTTGG AACTATTTAT TGAACAGTTG CTCTGTGTGA TGGATTTCGG GGATACCTAG
      ATGGAATGGG CATGATCCCC TTTTTACAGA AATAGAAAAT AGGTGGCCTA TGAATTATTC
      TTTCCTTTTA TATCCATGAC AAACTTTAGT AAAAAATTTT CTTTTCTACT GAGCTTAGCA
      TTTATTCAGT ACTCTTCTCA ATATATTTTC CAGGTAGCTA GTGACAATCA GAGTGATATG
35    TAAGACAAAC TCATTTGTCC TCCTAGTAGG AAAGATTCTA GTAGAAGCAA AGAATTGTGT
      ACCATTCTGC AAGTGGTTTG TTGGAATCTT TCTTTGATAC CTGTTCCTGT ATTCCCCAC
      CCCCATTAAT TTAGTCATTA ATTACTACAT GAACCCGTAA AATAAATCCT TAAATTATTT
      TCCTGGTAGA TTTTTTGAGC TGTGTAAGGA CCTTTCAATT CACTTTACAT TAGAATACGA
      TATGTGATGG TAAGTATTAA CCCAGCTTCC TGAGTGATGG CGTGAGGGCG GAACCCAGGT
40    TCATGAAAAG TATTCCTATT ATTGAATTTT ACCAGTTATT TTCAGAAGTT TGATACAATG
      TGAGTTGATT TCATAACATG TCCTCTAACT GCACTTGATA GCAAATTATC TTGTTATCCT
      GAGTTGTAGC CAACAATGAC TTGGAGAATC TATGCAATAC TCAGTTTTAT TACTTCTAAG
      CTCATTTTGA AGATAATACT ACCCATGGCT GATTTGTTAC TATAAAATAG GTTTAGTATT
      TGCTGTCTGG AAACATTCTG ATTAGTGTCT CCTGGGAGGA TTATAAATTT ATAGTACCCA
```

```
AAGANTAAAC NTGTTGTTTC CCTTTCCTAA ACTTTTAGTG NATAATNCAG TCTCTGCCGT
GTCTCATTTT CATCACTTGC CCTCCANAGC TCCCATCTCA CTGAATTCTT GCAGTGTTCT
GAATGTTGAG AGCCCCAANG TGGGTCTTAT AACAGCCAGT CAGCAACATT TCTGTTTTTC
ATCTGACACC AAGGGTCTCG TCTCTTTGCT TTTCTACCAG TTATTCTGGG CTCTTCAGCT
CTAAAGAAAG TATAGGTCCT GAAATCTTTC CCTACCTTCT CAATTTCCTG GGGAGGGCTT
CTTTGGAAAG TGGGATTGGA AATAAGATAA ATTTGAAGAT AATTGAGAAA TGAATGGAAA
GTGAAATTGA AGGGTCCATG TTAAGAGATT GCAAGTTATG CTATCACCAA ATAGATTTTT
TGTGCCTGAG AGGATTAATT CATAGTGCAT ATTATGTGTT GACTTATCA TTGAGGTCCT
GGCACATGAT AGCATTGGCA TGATATAATT TGAGCTACTG ATACTATAGT GTTGCTTCTG
GTGTTGTTAC CAGAACATCT AAAATATATT AGGATTTTTT TAATGGCAGA GGAAATGAAG
ACTAATATGA CATAGTCCTT GTCCTAGTGA TTTACAGTTT AGCAAGACAT ACAAGCAAAA
CATTAAAGTA AAGCATGATA ACTACTTTAA TAAAGCATTT TTTAAATTCA TTGGTAACAT
AAGAGAAGGT AGAAGAGTTT AGCAAACCCT TCCCAAAAGA AATGATTGAC AAGTTATATG
AGATAATAAT TCAGGGAAAG GAAATTCGAC TTTCTAAAGC CAAATTATTT GACATTGGTT
TTCATATAGC TTGGTAAAAG CTGTTATTTT CTCCCATGTT CTTTATTCTT TGACTGTTAT
AAGTATGATT TGTACAGAGA AATGGCAATT TCAAAACAGA GGGCTTTGAT GGATTAATTG
CTTTGAATTG ATCCCTCATC TACAGTATCT TGTCAGGTAC TTGAGAAAAT AATGTACTTA
AAGTTTCCTC TTTTGACTTT CTTTTGGTAT TCTATACTGT AAGTTGGGGA AAAAAGTATT
TTCTCTTCCT GCTAATTGGG CTACTTGAAA ATTCCCACCA ACTTTGCCAA TACCAGTGTT
CTGTATAACC CAGAATTCAG AATTAGCTGC AATTAAGGGA ATTCACAGCT TTTCTGTAGT
CAGAGAGCAA ATTGAAATTA AAGAAAAAAG AAATAGTGGG AGGACAAATG AGGTTTTACC
TTTACACTTG AAAACAGATT TAAGAACAAG CCTTATACCT AGATTTATTA ATACTTTGAG
GTATGAGAGG GAAGAGAAGC TTAGAAATAC GGCAGAATGG GCTTTCTTTG TTCTTCTCCC
AGCTATGCTG TTTTTATTTA TTATTGTATT TTTAAGAGAG AAGGGAAGTG TCTCTCCTGG
GTCACATTAA TTAGGAAATA CAGAGTGTTT TCATAATGCG TAAAGTCTAG TCCATTTAAG
TCTTGTTTCA AAATGCTATT TATATTATTT GAGCAGGAAG GCAGAGACCT TAAACTGCTC
CACCCAATTC ATTTTACACA AAAGATTAAA AAGAAAAAAA CAGTGTCAAA AGGTCAAGTG
CCCAGTGTTG CACAACTAAT GATAGGCAAA ACAAGAGAGG AGGGGAAAAA AAAGAATCCT
GACTCCCGAC TCCTAGTTCA GTGTGTTCTT CACTGTTAGA AGGTGCTGCT GAACATAGTT
ATACCATATG AAGATCACAC TATCTATTTG AGATTGTAGA AATTTGATTA CTGCAGAGCT
CTGGCTGGGC TAGATTCACT TCTTATTCTT TATTGCATTG CAGTATTCTT AAGAGATAAA
TGGCTCTTTT AGAATCAGAT TGGCCTTGGC TGTTGAAATG AGGCATTAAT TACCTTGGTA
GCTGACACAT TCTTACAGGT CAGGGCTTG ATGAAGTTTA TCTTCTTCCC TTGTTCCTGT
GATTGCTCTG TAAATAGAAC ACATTCAGAG CCCTTGAATG CACTTGCAGC CTGTGCCTCC
CACAGTGATC GATGGTCAGA TAATGGGAGT TTAATGACCA GTACTGAGAG AGATTATTTC
CATGGCTGCT ATGGGCCAAG AAAGCTGGGT GGTCAGAAAG GGACCTTTTC CAGACTCTCC
TGGGTTGTGT TATTTCTTTG ACATTGGTTT CCTTTCATGA GGCGCCAAGG TGTATTTGTA
AGTTGTCCAG TGTTGCACAG CTAATGAGGC GCAAACAAG AAAGGAGGAA AAAAAAGTA
TCCTCACTCC TAGTTCAGTG TGTTCTTCAC TGTTGGAAGG TACTGCTGAA CACAGTTATA
TGTTCAGTGT ACAATGTACA ATGTTCTTTG ACATTCCGTC ATTTGAAGGA TGGGTCCTAT
GGATAGTATC ATCTGCATGG TTTTGAAAAC AAAAGATATA AGCTAATTTT GCCCTGTCTA
GTGACTACGA GACAGGGAGA GAAAATCTGA ATATTTGTTA AAGTAGACAC AGACCCATAA
ATTGAAAAGG ACACTAATCC TGCCTTAGGA GACAGTAAGG CACTTGTCCC TGCTATCTAT
TAGCTGTGTG TCCTTGGACA GCTCATTGCT TCTTTCTGAG CCACCGTTAC ACGTTATCTA
```

34

```
TCAAGTGAGC AGGTTGTACA CTAGATGAAT TCACAGGTCC TTCTTCCAAG TGCTTTTCTA
ATCTTCATAA TTTAGATAAT CTCTCAGTAG CAAAACAGTG TACAATATGG TCAATCTGAG
ATTTTTAGGG GGAGAATTTT AGGGAACATC AGAAATGGCA GTAGTTAAAA GGAAATAGGA
CTCACAGGCT GACTTCTCTC TATAACTTCA CATGGTAGAA GGGATGAGAG TTCTCTCTGG
```
₅
```
AGCCTTTTGT ATTAGGTCAC TAATCCCAAA GGCCCCACCT CCTAAGACCA TCACTTTTGG
GATTGGGGTT TTAACATGAT TTTAGAGGGA CATAAACCTT CATCTCATTG CATTAGAGAT
ATTTGAAAGC TCAGCTCACG TGTATTCCTC CACAGCTCAC ATGTATTCCT CCATCACCCA
ACCTGATGGC TTTGAACGTT GTAGACATAA ATCCTTTCAT CATTATCAAG AATATTGCCA
AAAGCTTCTC AGAAATTATG AGGGGTTTTT TTAGTTTCTA AAATATTCTC AAAGAAAGTC
```
₁₀
```
CCATGTACTA ATGTTTGCCT TTTGATGAAA AAGGATGAAA TCTTAATGAT TGCCTTAATA
AGCTCAACAA TGCTTGTTAG TTGAGTCTTC TTATTGTGCT GATTCTTATA AACAACAACA
TTCAGTATAA ACATTAATGT ATGTGATTCA CTAAGGTTTT TGCATGATTC TCTGTGAGGC
TTCAGATGTC TCTTGGATTA TGTGTCTTTT TTTCATCGCC AGCATCCTGT GGCAGCTGCT
ACTCATTTGC TTCTATGGGT ATGCTAGAAG CGAGAATCCG TATACTAACC AACAATTCTC
```
₁₅
```
AGACCCCAAT CCTAAGCCCT CAGGAGGTTG TGTCTTGTAG CCAGTATGCT CAAGGTAAGT
GTTGCATTTC AGACACCATT TATGAGCTAT TTACCTGTGT GCAGCTGGCT GTTGTTGGCA
AAGGCAAAAG GATGATGCAG TAGAGAGAGC GCAGTGTCTA TAGTCAGAAA ATCTGAGTGC
AAGTCTGGCC CTATCACTTA TTAATGGATG ATTGCTCATG GAATTACTG TACCATCCAG
CAAAATGTCA ATAGTTACTA TATATTGAGT GAGCTCTGCT TGTTACATAT ATTGCCTAAC
```
₂₀
```
AATGCTCAAA ACTCTGAGAA GTAGTAAGTA TAATCCCTAT TTATGGGCGG GGAACAGGAA
CTAAGAAATT TTTCTAATAA TTTGAAGGTC TCACAGCTTT TAGCATTGGA GTTTCACTTC
TAATCATCGT CTCCAAAACC CAACTTTTAT TAAAACTATA CTAACACTGG TTTCTCTCTG
GGAGAATTTT AAAATTCTGT ACTTAGGGCT GGGCACAGTG GCTTATGCCT ATAACCCTAT
CACTTTGGGA GGCTGAGATG GGTGAATCTC TTGAGTCCTA GAGTTTGAGA CCAGCCTGGG
```
₂₅
```
CAACACGGCG AAACCCCTTC TCTATTAAAA ATACAAAAAA TTAGCTGGGC GTGGTGGTGT
GTGCTTGTAG TCCCAGCTAT TCAGGAGGCG GAGGTGTAAG AATCACCTGA GCCCAGGAGG
TCAAGGCTGC AGTGAGCCGA TATCATGCCA CTGCACTCCA GCCTGGGCAA ACGGAGTGAG
GCCCTGTTAT GAAAAAAAA AAATCTGTAC TTAGGCTTTC AGATCAGGCT GTATGTGATG
TATGTCGAAA ACACAGCTAT AATTGATTGA GGGAGAAACG TTACCATTTT AAAGTTTATG
```
₃₀
```
CTTTCAAGCC CAGATTTGGC CACTAGGAAT TTCCCAGCTC ACTAGTGAAA CTGCTGATGA
GTGATTATTT GCCAGTGAGC CTTTCATTCT TTCTAAAATA TGTACTACTA GTTGTGACTT
GTAGGCTATA GGGGCTATAA TATATCAAGA CAATCTTTAT CCTCATGAAG CTTACAGTTA
AGTAAGAGAT AGAGATTAAA TAATTATAAC AACAGAGTGA AGAACAGTGA AGAAAAAGTA
CAGAGTTATA ATATATATAA TAGGGCCAGG ACTGCATGAG GAAGGTAGGA AAGACATTTC
```
₃₅
```
GGCAAGAGGT TGTCAGGGAA AAGACTTGCT TGAGAAAGAG CCAAGTTGTG GGGTCTGGCT
GCTTAGCAAT GACCATAATA CCTAACTTTT GCTATTTTTA CATGAAGTAA CTAATTTAAC
CCTATGAGGA AAGTACTACT ACCATCTAGA TTTTACAGGT AAGTAAGCAG AGATACAGAG
AAGTTAAACT CTTCACACGG CTTTGGCTTT AAACCTATAT AGGCTTCAGA GCCTCCCCAC
TTAACCACTT TGCCATAGCT ACATCCATAT TAGGTGCTAA GTAGATATCT GTTAAGTAGA
```
₄₀
```
AGGAGGATGA AAGGATAGTT AGCTAGTTGG AAAATGGATG GATGAATGAA GTGATGCTTA
AGCTAAGAAC AACTTTCAGG GGTAACATGC AAAGAATAAT GGAGCAAAGA AGAAAAAATA
GAAAATGGGA TAATCCTTTT TCTACTAAGG GGTAACCATG TGTGTTATTC ATCTTCAGGC
TGTGAAGGCG GCTTCCCATA CCTTATTGCA GGAAAGTACG CCCAAGATTT TGGGCTGGTG
GAAGAAGCTT GCTTCCCCTA CACAGGCACT GATTCTCCAT GCAAAATGAA GGAAGACTGC
```
₃₅

TTTCGTTATT ACTCCTCTGA GTACCACTAT GTAGGAGGTT TCTATGGAGG CTGCAATGAA

GCCCTGATGA AGCTTGAGTT GGTCCATCAT GGGCCCATGG CAGTTGCTTT TGAAGTATAT

GATGACTTCC TCCACTACAA AAAGGGGATC TACCACCACA CTGGTCTAAG AGACCCTTTC

AACCCCTTTG AGCTGACTAA TCATGCTGTT CTGCTTGTGG GCTATGGCAC TGACTCAGCC

5   TCTGGGATGG ATTACTGGAT TGTTAAAAAC AGCTGGGGCA CCGGCTGGGG TGAGAATGGC

TACTTCCGGA TCCGCAGAGG AACTGATGAG TGTGCAATTG AGAGCATAGC AGTGGCAGCC

ACACCAATTC CTAAATTGTA GGGTATGCCT TCCAGTATTT CATAATGATC TGCATCAGTT

GTAAAGGGGA ATTGGTATAT TCACAGACTG TAGACTTTCA GCAGCAATCT CAGAAGCTTA

CAAATAGATT TCCATGAAGA TATTTGTCTT CAGAATTAAA ACTGCCCTTA ATTTTAATAT

10   ACCTTTCAAT CGGCCACTGG CCATTTTTTT CTAAGTATTC AATTAAGTGG GAATTTTCTG

GAAGATGGTC AGCTATGAAG TAATAGAGTT TGCTTAATCA TTTGTAATTC AAACATGCTA

TATTTTTTAA AATCAATGTG AAAACATAGA CTTATTTTTA AATTGTACCA ATCACAAGAA

AATAATGGCA ATAATTATCA AAACTTTTAA AATAGATGCT CATATTTTTA AAATAAAGTT

TTAAAA

15

The corresponding cDNA sequence for CTSC is

provided below (SEQ ID NO: 2):

          1 aattcttcac ctctttctc agctccctgc agcatgggtg ctgggccctc cttgctgctc

20   61 gccgccctcc tgctgcttct ctccggcgac ggcgccgtgc gctgcgacac acctgccaac

        121 tgcacctatc ttgacctgct gggcacctgg gtcttccagg tgggctccag cggttcccag

        181 cgcgatgtca actgctcggt tatgggacca caagaaaaaa aagtagtggt gtaccttcag

        241 aagctggata cagcatatga tgaccttggc aattctggcc atttcaccat catttacaac

        301 caaggctttg agattgtgtt gaatgactac aagtggtttg ccttttttaa gtataaagaa

25   361 gagggcagca aggtgaccac ttactgcaac gagacaatga ctgggtgggt gcatgatgtg

        421 ttgggccgga actgggcttg tttcaccgga aagaaggtgg gaactgcctc tgagaatgtg

        481 tatgtcaaca cagcacacct taagaattct caggaaaagt attctaatag gctctacaag

        541 tatgatcaca actttgtgaa agctatcaat gccattcaga gtcttggac tgcaactaca

        601 tacatggaat atgagactct taccctggga gatatgatta ggagaagtgg tggccacagt

30   661 cgaaaaatcc caaggcccaa acctgcacca ctgactgctg aaatacagca aaagattttg

        721 catttgccaa catcttggga ctggagaaat gttcatggta tcaattttgt cagtcctgtt

        781 cgaaaccaag catcctgtgg cagctgctac tcatttgctt ctatgggtat gctagaagcg

        841 agaatccgta tactaaccaa caattctcag accccaatcc taagccctca ggaggttgtg

        901 tcttgtagcc agtatgctca aggctgtgaa ggcggcttcc cataccttat tgcaggaaag

35   961 tacgcccaag attttgggct ggtggaagaa gcttgcttcc cctacacagg cactgattct

       1021 ccatgcaaaa tgaaggaaga ctgctttcgt tattactcct ctgagtacca ctatgtagga

       1081 ggtttctatg gaggctgcaa tgaagccctg atgaagcttg agttggtcca tcatgggccc

       1141 atggcagttg cttttgaagt atatgatgac ttcctccact acaaaaaggg gatctaccac

       1201 cacactggtc taagagaccc tttcaacccc tttgagctga ctaatcatgc tgttctgctt

40   1261 gtgggctatg gcactgactc agcctctggg atggattact ggattgttaa aaacagctgg

       1321 ggccaccggct ggggtgagaa tggctacttc cggatccgca gaggaactga tgagtgtgca

       1381 attgagagca tagcagtggc agccacacca attcctaaat tgtagggtat gccttccagt

       1441 atttcataat gatctgcatc agttgtaaag gggaattggt atattcacag actgtagact

                                    36

```
1501 ttcagcagca atctcagaag cttacaaata gatttccatg aagatatttg tcttcagaat
1561 taaaactgcc cttaatttta atataccttt caatcggcca ctggccattt ttttctaagt
1621 attcaattaa gtgggaattt tctggaagat ggtcagctat gaagtaatag agtttgctta
1681 atcatttgta attcaaacat gctatatttt ttaaaatcaa tgtgaaaaca tagacttatt
1741 tttaaattgt accaatcaca agaaaataat ggcaataatt atcaaaactt ttaaaataga
1801 tgctcatatt tttaaaataa agttttaaaa ataactgc
```

The wild type CTSC protein sequence is set forth below as SEQ
ID NO: 3:

MGAGPSLLLAALLLLLSGDGAVRCDTPANCTYLDLLGTWVFQVG
SSGSQRDVNCSVMGPQEKKVVVYLQKLDTAYDDLGNSGHFTIIYNQGFEIVLNDYKWF
AFFKYKEEGSKVTTYCNETMTGWVHDVLGRNWACFTGKKVGTASENVYVNTAHLKNSQ
EKYSNRLYKYDHNFVKAINAIQKSWTATTYMEYETLTLGDMIRRSGGHSRKIPRPKPA
PLTAEIQQKILHLPTSWDWRNVHGINFVSPVRNQASCGSCYSFASMGMLEARIRILTN
NSQTPILSPQEVVSCSQYAQGCEGGFPYLIAGKYAQDFGLVEEACFPYTGTDSPCKMK
EDCFRYYSSEYHYVGGFYGGCNEALMKLELVHHGPMAVAFEVYDDFLHYKKGIYHHTG
LRDPFNPFELTNHAVLLVGYGTDSASGMDYWIVKNSWGTGWGENGYFRIRRGTDECAI
ESIAVAATPIPKL.

Papillon Lefevre syndrome (PLS) is an autosomal
recessive disorder characterized by palmoplantar
hyperkeratosis and severe early onset periodontitis that
results in the premature loss of the primary and
secondary dentitions. The 46 kb CTSC gene consists of 7
exons and is mutated in PLS patients. Sequence analysis
of CTSC from PLS affected individuals from thirty-two
Turkish families identified four different mutations.
An exon 6 nonsense mutation (856C->T) introduces a
premature stop codon at amino acid 286. Three exon 2
mutations were identified including a single nucleotide
deletion (1047delA) of codon 349 introducing a
frameshift and premature termination codon, a two base
pair deletion (1028-1029delCT) that results in
introduction of a stop codon at amino acid 343, and a
G->A substitution in codon 429 (1286G->A) introducing a
premature termination codon. All PLS affected
individuals examined were homozygous for cathepsin C
mutations inherited from a common ancestor. Parents and
siblings heterozygous for cathepsin C mutations do not
show either the palmoplantar hyperkeratosis or severe
early onset periodontitis characteristic of PLS. In

37

addition to the 5 families described above, Table I
summarizes CTSC mutations identified in 27 other
families presenting with symptoms of PLS.

5   Haim-Munk syndrome is a rare condition associated
with congenital palmoplantar keratosis, pes planus,
onychogyrphosis, periodontosis, arachnodactyly and
acroosteolysis. In an additional embodiment of the
invention, a mutation in cathepsin C causing Haim-Munk
Syndrome has been identified. It appears that

10   substitution of an A for a G at CTSC nucleotide position
857 in Exon 6 is responsible for this syndrome in
patients.

            Based on the data presented herein, it appears that
additional mutations or functional polymorphisms are

15   associated with other pathological conditions,
including, but not limited to prepubertal periodontitis
(PPP), early onset periodontal disease or other forms of
gum disease. For example, as shown herein, PPP is
caused a substitution of a G for an A at position 1040

20   in the CTSC coding sequence. Thus, the invention also
provides methods for screening the CTSC gene for
alterations associated with these disease states.

      I.   **Preparation of Altered Human CTSC-Encoding Nucleic**

25   **Acid Molecules, CTSC Proteins, and Antibodies**

             Thereto

             A.   **Nucleic Acid Molecules**

                  Nucleic acid molecules encoding the human CTSC
proteins of the invention may be prepared by two general

30   methods: (1) synthesis from appropriate nucleotide
triphosphates, or (2) isolation from biological sources.
Both methods utilize protocols well known in the art.
The availability of nucleotide sequence information,
such as a DNA having the sequence of SEQ ID NOS:1-2

35   enables preparation of an isolated nucleic acid molecule
of the invention by oligonucleotide synthesis.
Synthetic oligonucleotides may be prepared by the

phosphoramidite method employed in the Applied
Biosystems 38A DNA Synthesizer or similar devices. The
resultant construct may be purified according to methods
known in the art, such as high performance liquid
5   chromatography (HPLC). Long, double-stranded
polynucleotides, such as a DNA molecule of the present
invention, must be synthesized in stages, due to the
size limitations inherent in current oligonucleotide
synthetic methods. Thus, for example, a 4.7 kb double-
10   stranded molecule may be synthesized as several smaller
segments of appropriate complementarity. Complementary
segments thus produced may be annealed such that each
segment possesses appropriate cohesive termini for
attachment of an adjacent segment. Adjacent segments
15   may be ligated by annealing cohesive termini in the
presence of DNA ligase to construct an entire 4.7 kb
double-stranded molecule. A synthetic DNA molecule so
constructed may then be cloned and amplified in an
appropriate vector.

20   Nucleic acid sequences encoding the altered human
CTSC proteins of the invention may be isolated from
appropriate biological sources using methods known in
the art. In a preferred embodiment, a cDNA clone is
isolated from a cDNA expression library of human origin.
25   In an alternative embodiment, utilizing the sequence
information provided by the cDNA sequence, human genomic
clones encoding altered CTSC proteins may be isolated.

        Table 1 sets forth several different mutations
associated with particular PPKs and PPP. Altered CTSC-
30   specific probes for identifying such sequences may be
between 15 and 40 nucleotides in length. For probes
longer than those shown above, the additional contiguous
nucleotides are provided within SEQ ID NOS:1 and 2.

        Additionally, cDNA or genomic clones having
35   homology with human CTSC may be isolated from other
species using oligonucleotide probes corresponding to
predetermined sequences within the human CTSC encoding

nucleic acids.

In accordance with the present invention, nucleic
acids having the appropriate level of sequence homology
with the protein coding region of SEQ ID NO:1 may be
identified by using hybridization and washing conditions
of appropriate stringency. For example, hybridizations
may be performed, according to the method of Sambrook et
al., Molecular Cloning, Cold Spring Harbor Laboratory
(1989), using a hybridization solution comprising: 5X
SSC, 5X Denhardt's reagent, 1.0% SDS, 100 µg/ml
denatured, fragmented salmon sperm DNA, 0.05% sodium
pyrophosphate and up to 50% formamide. Hybridization is
carried out at 37-42°C for at least six hours.
Following hybridization, filters are washed as follows:
(1) 5 minutes at room temperature in 2X SSC and 1% SDS;
(2) 15 minutes at room temperature in 2X SSC and 0.1%
SDS; (3) 30 minutes-1 hour at 37°C in 1X SSC and 1% SDS;
(4) 2 hours at 42-65°C in 1X SSC and 1% SDS, changing
the solution every 30 minutes.

Nucleic acids of the present invention may be
maintained as DNA in any convenient cloning vector. In
a preferred embodiment, clones are maintained in a
plasmid cloning/expression vector, such as pBluescript
(Stratagene, La Jolla, CA), which is propagated in a
suitable *E. coli* host cell.

Altered CTSC-encoding nucleic acid molecules of the
invention include cDNA, genomic DNA, RNA, and fragments
thereof which may be single- or double-stranded. Thus,
this invention provides oligonucleotides having
sequences capable of hybridizing with at least one
sequence of a nucleic acid molecule of the present
invention, such as selected segments of the DNA having
SEQ ID NO:1. Also contemplated in the scope of the
present invention are oligonucleotide probes which
specifically hybridize with the mutated CTSC genes of
the invention while not hybridizing with the wild type
sequence under high stringency conditions. Primers

4C

capable of specifically amplifying the altered CTSC
encoding nucleic acids described herein are also
contemplated herein. As mentioned previously, such
oligonucleotides are useful as probes and primers for
5   detecting, isolating or amplifying altered CTSC genes.

Antisense nucleic acid molecules may be targeted to
translation initiation sites and/or splice sites to
inhibit the expression of the CTSC gene or production of
the CTSC protein of the invention. Such antisense
10   molecules are typically between 15 and 30 nucleotides in
length and often span the translational start site of
CTSC encoding mRNA molecules.

Alternatively, antisense constructs may be
generated which contain the entire CTSC cDNA in reverse
15   orientation. Such antisense constructs are easily
prepared by one of ordinary skill in the art.

It will be appreciated by persons skilled in the
art that variants (e.g., allelic variants) of CTSC
sequences exist in the human population, and must be
20   taken into account when designing and/or utilizing
oligonucleotides of the invention. Accordingly, it is
within the scope of the present invention to encompass
such variants, with respect to the CTSC sequences
disclosed herein or the oligonucleotides targeted to
25   specific locations on the respective genes or RNA
transcripts. Accordingly, the term "natural allelic
variants" is used herein to refer to various specific
nucleotide sequences of the invention and variants
thereof that would occur in a human population. The
30   usage of different wobble codons and genetic
polymorphisms which give rise to conservative or neutral
amino acid substitutions in the encoded protein are
examples of such variants. Such variants would not
demonstrate altered CTSC activity. Additionally, the
35   term "substantially complementary" refers to
oligonucleotide sequences that may not be perfectly
matched to a target sequence, but such mismatches do not

WO 01/07663

PCT/US00/20400

materially affect the ability of the oligonucleotide to hybridize with its target sequence under the conditions described.

5

B. Proteins

Full-length, altered, human CTSC proteins of the present invention may be prepared in a variety of ways, according to known methods. The proteins may be

10   purified from appropriate sources, e.g., transformed bacterial or animal cultured cells or tissues, by immunoaffinity purification. However, this is not a preferred method due to the low amount of protein likely to be present in a given cell type at any time. The

15   availability of nucleic acid molecules encoding CTSC protein enables production of the protein using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into an appropriate *in vitro* transcription vector, such as pSP64 or pSP65 for *in*

20   *vitro* transcription, followed by cell-free translation in a suitable cell-free translation system, such as wheat germ or rabbit reticulocyte lysates. *In vitro* transcription and translation systems are commercially available, e.g., from Promega Biotech, Madison,

25   Wisconsin or Gibco-BRL, Gaithersburg, Maryland.

Alternatively, according to a preferred embodiment, larger quantities of CTSC protein may be produced by expression in a suitable prokaryotic or eukaryotic system. For example, part or all of a DNA molecule,

30   such as a DNA having SEQ ID NOS:1 or 2 containing an alteration set forth in Table 1 may be inserted into a plasmid vector adapted for expression in a bacterial cell, such as *E. coli*. Such vectors comprise the regulatory elements necessary for expression of the DNA

35   in the host cell positioned in such a manner as to permit expression of the DNA in the host cell. Such regulatory elements required for expression include

42

promoter sequences, transcription initiation sequences and, optionally, enhancer sequences.

The human CTSC protein produced by gene expression in a recombinant procaryotic or eukaryotic system may be
5 purified according to methods known in the art. In a preferred embodiment, a commercially available expression/secretion system can be used, whereby the recombinant protein is expressed and thereafter secreted from the host cell, and readily purified from the
10 surrounding medium. If expression/secretion vectors are not used, an alternative approach involves purifying the recombinant protein by affinity separation, such as by immunological interaction with antibodies that bind specifically to the recombinant protein or nickel
15 columns for isolation of recombinant proteins tagged with 6-8 histidine residues at their N-terminus or C-terminus. Alternative tags may comprise the FLAG epitope or the hemagglutinin epitope. Such methods are commonly used by skilled practitioners.
20 The human CTSC protein of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures. For example, such protein may be subjected to amino acid sequence analysis, according to known methods.
25 The present invention also provides antibodies capable of immunospecifically binding to proteins of the invention. Polyclonal antibodies directed toward altered human CTSC proteins may be prepared according to standard methods. In a preferred embodiment, monoclonal
30 antibodies are prepared, which react immunospecifically with the various epitopes of the CTSC protein described herein. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. Polyclonal or monoclonal antibodies
35 that immunospecifically interact with altered CTSC proteins can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized

43

for affinity separation of proteins with which they
immunospecifically interact. Antibodies may also be
used to immunoprecipitate proteins from a sample
containing a mixture of proteins and other biological

5   molecules. Other uses of anti-CTSC antibodies are
described below.

II.   **DETECTION OF KERATODERMAL DISORDERS/DYSPLASIAS and**
      PERIODONTAL DISEASE-ASSOCIATED MUTATIONS AND
10   **DIAGNOSTIC SCREENING ASSAYS THEREFORE**

Currently, the most direct method for mutational
analysis is DNA sequencing, however it is also the most
labor intensive and expensive. It is usually not

15   practical to sequence all potentially relevant regions
of every experimental sample. Instead some type of
preliminary screening method is commonly used to
identify and target for sequencing only those samples
that contain mutations. Single stranded conformational

20   polymorphism (SSCP) is a widely used screening method
based on mobility differences between single-stranded
wild type and mutant sequences on native polyacrylamide
gels. Other methods are based on mobility differences
in wild type/mutant heteroduplexes (compared to control

25   homoduplexes) on native gels (heteroduplex analysis) or
denaturing gels (denaturing gradient gel
electrophoresis). Sample preparation is relatively easy
in these assays, and conditions for electrophoresis
required to generate the often subtle mobility

30   differences that form the basis for identifying the
targets that contain mutations are known to those of
skill in the art. Another parameter to be considered is
the size of the target region being screened. In
general, SSCP is used to screen target regions no longer

35   than about 200-300 bases.
Another type of screening technique currently in
use is based on cleavage of unpaired bases in

44

heteroduplexes formed between wild type probes
hybridized to experimental targets containing point
mutations. The cleavage products are also analyzed by
gel electrophoresis, as subfragments generated by
cleavage of the probe at a mismatch generally differ
significantly in size from full length, uncleaved probe
and are easily detected with a standard gel system.
Mismatch cleavage has been effected either chemically
(osmium tetroxide, hydroxylamine) or with a less toxic,
enzymatic alternative, using RNase A. The RNase A
cleavage assay has also been used, although much less
frequently, to screen for mutations in endogenous mRNA
targets for detecting mutations in DNA targets amplified
by PCR. A mutation detection rate of over 50% was
reported for the original RNase screening method.

A newer method to detect mutations in DNA relies on
DNA ligase which covalently joins two adjacent
oligonucleotides which are hybridized on a complementary
target nucleic acid. The mismatch must occur at the
site of ligation. As with other methods that rely on
oligonucleotides, salt concentration and temperature at
hybridization are crucial. Another consideration is the
amount of enzyme added relative to the DNA
concentration. In summary, exemplary approaches for
detecting alterations in CTSC encoding nucleic acids or
polypeptides/proteins include:

a) comparing the sequence of nucleic acid in the
sample with the wild-type CTSC nucleic acid sequence to
determine whether the sample from the patient contains
mutations; or

b) determining the presence, in a sample from a
patient, of the polypeptide encoded by the CTSC gene
and, if present, determining whether the polypeptide is
full length, and/or is mutated, and/or is expressed at
the normal level; or

c) using DNA restriction mapping to compare the
restriction pattern produced when a restriction enzyme

cuts a sample of nucleic acid from the patient with the restriction pattern obtained from normal CTSC gene or from known mutations thereof; or,

d) using a specific binding member capable of binding to a CTSC nucleic acid sequence (either normal sequence or known mutated sequence), the specific binding member comprising nucleic acid hybridizable with the CTSC sequence, or substances comprising an antibody domain with specificity for a native or mutated CTSC nucleic acid sequence or the polypeptide encoded by it, the specific binding member being labeled so that binding of the specific binding member to its binding partner is detectable; or,

e) using PCR involving one or more primers based on normal or mutated CTSC gene sequence to screen for normal or mutant CTSC gene in a sample from a patient.

A "specific binding pair" comprises a specific binding member (sbm) and a binding partner (bp) which have a particular specificity for each other and which in normal conditions bind to each other in preference to other molecules. Examples of specific binding pairs are antigens and antibodies, ligands and receptors and complementary nucleotide sequences. The skilled person is aware of many other examples and they do not need to be listed here. Further, the term "specific binding pair" is also applicable where either or both of the specific binding member and the binding partner comprise a part of a large molecule. In embodiments in which the specific binding pair are nucleic acid sequences, they will be of a length to hybridize to each other under conditions of the assay, preferably greater than 10 nucleotides long, more preferably greater than 15 or 20 nucleotides long.

In most embodiments for screening for susceptibility alleles, the CTSC nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other

46

sequences present in the sample. This allows the target
CTSC sequences to be detected with a high degree of
sensitivity if they are present in the sample. This
initial step may be avoided by using highly sensitive
array techniques that are becoming increasingly
important in the art.

The identification of the CTSC gene and its
association with keratodermal disorders/dysplasias and
peridontal diseases paves the way for aspects of the
present invention to provide the use of materials and
methods, such as are disclosed and discussed above, for
establishing the presence or absence in a test sample of
a variant form of the gene, in particular an allele or
variant specifically associated with PLS, HMS or
periodontal diseases. This may be for diagnosing a
predisposition of an individual to PLS, HMS or
periodontal disease. It may be for diagnosing PLS, HMS
or periodontal disease in a patient with the disease as
being associated with the altered CTSC gene.

This allows for planning of appropriate therapeutic
and/or prophylactic measures, permitting stream-lining
of diagnosis, treatment and outcome assessments. The
approach further stream-lines treatment by targeting
those patients most likely to benefit.

According to another aspect of the invention,
methods of screening drugs for therapy to identify
suitable drugs for restoring CTSC product functions are
provided.

The CTSC polypeptide or fragment employed in drug
screening assays may either be free in solution, such as
gingival crevicular fluid, affixed to a solid support or
within a cell. One method of drug screening utilizes
eukaryotic or prokaryotic host cells which are stably
transformed with recombinant polynucleotides expressing
the polypeptide or fragment, preferably in competitive
binding assays. Such cells, either in viable or fixed
form, can be used for standard binding assays. One may

determine, for example, formation of complexes between a
CTSC polypeptide or fragment and the agent being tested,
or examine the degree to which the formation of a
complex between a CTSC polypeptide or fragment and a
5   known ligand is interfered with by the agent being
tested.

Another technique for drug screening provides high
throughput screening for compounds having suitable
binding affinity to CTSC polypeptides and is described
10   in detail in Geysen, PCT published application WO
84/03564, published on Sep. 13, 1984. Briefly stated,
large numbers of different, small peptide test compounds
are synthesized on a solid substrate, such as plastic
pins or some other surface. The peptide test compounds
15   are reacted with CTSC polypeptide and washed. Bound
CTSC polypeptide is then detected by methods well known
in the art.

A further technique for drug screening involves the
use of host eukaryotic cell lines or cells (such as
20   described above) which have a nonfunctional CTSC gene.
These host cell lines or cells are defective at the CTSC
polypeptide level. The host cell lines or cells are
grown in the presence of drug compound. The rate of
growth of the host cells is measured to determine if the
25   compound is capable of regulating the growth of CTSC
defective cells.

The goal of rational drug design is to produce
structural analogs of biologically active polypeptides
of interest or of small molecules with which they
30   interact (e.g., agonists, antagonists, inhibitors) in
order to fashion drugs which are, for example, more
active or stable forms of the polypeptide, or which,
e.g., enhance or interfere with the function of a
polypeptide in vivo. See, e.g., Hodgson, (1991)
35   Bio/Technology 9:19-21. In one approach, one first
determines the three-dimensional structure of a protein
of interest (e.g., CTSC polypeptide) or, for example,

4 8

of the CTSC-substrate complex, by x-ray crystallography, by nuclear magnetic resonance, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a
5   polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., (1990) Science 249:527-533). In addition, peptides (e.g., CTSC polypeptide)
10   may be analyzed by an alanine scan (Wells, 1991) Meth. Enzym. 202:390-411. In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner
15   to determine the important regions of the peptide.

     It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug
20   design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be
25   an analog of the original molecule. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

30   Thus, one may design drugs which have, e.g., improved CTSC polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of CTSC polypeptide activity. By virtue of the availability of cloned CTSC sequences, sufficient amounts of the CTSC
35   polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the CTSC protein sequence

provided herein will guide those employing computer
modeling techniques in place of, or in addition to x-ray
crystallography.

5

III Therapeutics

A.   Pharmaceuticals and Peptide Therapies

The discovery that mutations in the CTSC gene give
rise to PLS, HMS, and periodontal disease facilitates
10   the development of pharmaceutical compositions useful
for treatment and diagnosis of these syndromes and
conditions. These compositions may comprise, in
addition to one of the above substances, a
pharmaceutcally acceptable excipient, carrier, buffer,
15   stabilizer or other materials well known to those
skilled in the art. Such materials should be non-toxic
and should not interfere with the efficacy of the active
ingredient. The precise nature of the carrier or other
material may depend on the route of administration, e.g.
20   oral, intravenous, cutaneous or subcutaneous, nasal,
intramuscular, intraperitoneal routes.

Whether it is a polypeptide, antibody, peptide,
nucleic acid molecule, small molecule or other
pharmaceutically useful compound according to the
25   present invention that is to be given to an individual,
administration is preferably in a "prophylactically
effective amount" or a "therapeutically effective
amount" (as the case may be, although prophylaxis may be
considered therapy), this being sufficient to show
30   benefit to the individual.

B.   Methods of Gene Therapy

As a further alternative, the nucleic acid encoding
the authentic biologically active CTSC polypeptide could
35   be used in a method of gene therapy, to treat a patient
who is unable to synthesize the active "normal"
polypeptide or unable to synthesize it at the normal

level, thereby providing the effect elicited by wild-
type CTSC and suppressing the occurrence of "abnormal"
CTSC associated with keratodermal disorders and
dysplasias.

5   Vectors, such as viral vectors have been used in
the prior art to introduce genes into a wide variety of
different target cells. Typically the vectors are
exposed to the target cells so that transformation can
take place in a sufficient proportion of the cells to

10   provide a useful therapeutic or prophylactic effect from
the expression of the desired polypeptide. The
transfected nucleic acid may be permanently incorporated
into the genome of each of the targeted cells, providing
long lasting effect, or alternatively the treatment may

15   have to be repeated periodically.
A variety of vectors, both viral vectors and
plasmid vectors are known in the art, see US Patent No.
5,252,479 and WO 93/07282. In particular, a number of
viruses have been used as gene transfer vectors,

20   including papovaviruses, such as SV40, vaccinia virus,
herpes viruses including HSV and EBV, and retroviruses.
Many gene therapy protocols in the prior art have
employed disabled murine retroviruses.
Gene transfer techniques which selectively target

25   the CTSC nucleic acid to oral tissues are preferred.
Examples of this include receptor-mediated gene
transfer, in which the nucleic acid is linked to a
protein ligand via polylysine, with the ligand being
specific for a receptor present on the surface of the

30   target cells.

The following methods are provided to facilitate the
practice of the present invention.

35   **Family material and clinical diagnosis.**
Five Turkish families were described previously
[8]. All available family members provided consent for

the study and were clinically examined. A diagnosis of
PLS was made in individuals with severe early onset
periodontitis and the clinical appearance of
hyperkeratosis on the palmar and plantar surfaces. All
5   affected individuals also had hyperkeratosis on the
knees. DNA was isolated from peripheral blood samples
from all available members from these nuclear families
using standard techniques (Qiamp Blood Kit, Qiagen).

10   **RNA Isolation, Amplification, and Tissue Expression
Analysis.**
         Total RNA was generated from fresh tissue samples
(gingiva, palm, sole, knee) using TRIZOL reagent
(Molecular Research Center, Inc.; Cincinnati, OH)
15   according to the manufacturer's protocol. To determine
if cathepsin C was expressed in a given tissue, single-
tube RT-PCR was carried out using the Access RT-PCR
System (Promega; Madison, WI), following the
manufacturer's protocol. A portion of each reaction was
20   visualized following agarose gel electrophoresis in the
presence of ethidium bromide. Amplification primers
located within exon 6 **F** 5'-AGGAGGTTGTGTCTTGTAGCC-3'(nt.
857-877;SEQ ID NO: 4) and exon 7 **R** 5'-
AGTGCCTGTGTAGGGGAAGC-3' (NT 981-962; SEQ ID NO: 5)
25   produce an amplicon of 123 base pairs from cDNA. A
standard PCR protocol was followed with an annealing
temperature of 65°C.

GenBank accession numbers.
30   Full-length cDNA of CTSC (NM-001814) and full-
length genomic DNA of CTSC contained within a BAC
vector, Genbank accession number (AC011088). See SEQ ID
NO: 1.

35   **Cathepsin C Activity Assay**
         In unaffected non-carriers, cathepsin C activity
ranges from 600-1200 μmol/min/ mg. As carriers of a

cathepsin C mutation do not have clinical manifestations, measurement of cathepsin C enzymatic activity can be used to determine whether at-risk family members are carriers. Enzymatic activity can also be

5 used to determine if individuals marrying into a family are carriers. Carriers typically have approximately 50% of normal enzyme activity. Determination of enzymatic activity can also be used to establish a diagnosis of PLS when mutational studies of cathepsin C have been

10 negative. This is important in assuring that a diagnosis of PLS has been properly given to an individual with clinical symptoms suggestive of PLS.

Viable leukocyte pellets are obtained from lithium heparinized whole blood by mixing blood with 3 volumes

15 of 3% dextran in normal saline, and allowing the red cells to settle for 45 min at room temperature. Cells are pelleted by centrifugation at 1500 rpm for 5 min at 4*C. After washing and removal of contaminating red cells, leukocyte pellets are resuspended in dH2O and

20 sonicated on ice for 5 sec each for total of 6 blasts using a Sonic 300 Dismembrator. Protein concentration is determined by the Lowry method.

Enzymatic activity is determined by measuring hydrolysis of the synthetic substrate

25 glycyl-L-arginine-7-amido-4-methylcoumarin at a final concentration of 5 mM using a modified method. All reactions are performed in duplicate. Twenty µl of leukocyte lysate are added to 200 µl of Na_3PO_4 buffer (0.1M, pH 6.5) in a 96 well plate and then

30 substrate added. Reactions are allowed to proceed for 1 hr at room temperature at which time 10 µl of glycine-NaOH buffer (0.5M, pH 9.8) is added to stop the reaction. Fluorescence is determined using a Perkin-Elmer LS50B luminescence spectrometer at 370-nm

35 excitation and 460-nm emission. The amount of NHMec released is determined by generating a standard curve using NHMec. Cathepsin C activity is reported as µmol

NHMec released per min per mg protein.

Sequencing and mutation analysis.

   PCR primers were designed to cover the entire
5   cathepsin C gene in overlapping fragments, from 955
nucleotides 5' to the start codon to 240 nucleotides 3'
to the termination codon using cathepsin C (DPP-I)
sequence data (Accession # U79415; SEQ ID NO: 1). The
PCR products were prepared for sequencing by excising
10   the bands from the agarose gel and extracting the
fragments using a Qiagen Gel Clean-up Kit.   The sense
and antisense strand of each PCR product were directly
sequenced on an ABI Prism 310 Genetic Analyzer (Perkin-
Elmer) using four dye terminator chemistry.
15   Approximately 1-3 ng of purified product and 3.2 pmol
primer were added to premixed reagents from the ABI
Prism Big Dye Terminator Cycle Sequencing Ready Reaction
Kit, FS (Perkin-Elmer) and underwent a cycle sequencing
reaction in a GeneAmp PCR System 9700 (Perkin Elmer).
20   The linear amplification started with a 10 s
denaturation at 96°C, 5 s annealing at 50°C and 4 min
extension at 60°C.   The fluorescently labeled sequencing
products were separated from residual reaction reagents
using a Centri-Sep spin column (Princeton Separations,
25   Aldelphia NJ) and electrophoresed on POP6 capillary at
1500 V for 30 min.   Sequencing data were automatically
collected and analyzed by the ABI Prism 310 software.

Table A:

Primers Used to Determine Genomic Organization of
30   **Cathepsin C**

Region	Primer Sequences
Intron 1	F: 5'- TGTCAACTGCTCGGTTATGGGTAA-3' (#6) R: 5'- TCGAGCTTCTCTTCGTACACCACT-3' (#7)
Intron 2	F: 5'- TGACTACAAGTGGTTTGCCTTTTT-3' (#8) R: 5' TGCTGCCCTCTTCTTTATACTGC-3' (#9)
Intron 3	F: 5'- GCCTCTGAGAATGTGTATGTCAAC-3' (#10) R: 5' CCTGCCCCAAAAATGAGATA-3" (#11)

Intron 4	F: 5' TCGAAAAATCCCAAGGTAATC-3'	(#12)
	R: 5' GGGCCTAGAAAGGAAATATACATT-3'	(#13)
Intron 5	F: 5' AATTTGTTCGGAACTATTTATTGA-3'	(#14)
	R: 5' TCGCTTCTAGCATACCCATA-3'	(#15)

Exon	Primer Sequences	
1	F: 5'-GGCGATCAGACTGGCACACTAGAA -3'	(#16)
	R: 5'-CTTACCCATAACCGAGCAGTTGAC-3'	(#17)
2	F: 5'- GCAGACTGTGCTCAAACTGGGTAG-3'	(#18)
	R: 5'-TCTACTAATCAGAAGAGGTTTCAG-3'	(#19)
3	F: 5'-GGCACATTTACTGTGAATGAGAG-3'	(#20)
	R: 5'-GTCTCATTTGTAGCAACTCAC-3'	(#21)
4	F: 5'-CCACTTTCCACTTAGGCACAG-3'	(#22)
	R: 5'-AGGATGGTATTCAGCATTCATA-3'	(#23)
5	F: 5'-ATCCTAGCTAGTCTGGTAGCTGAA-3'	(#24)
	R: 5'-TCTAGGTATCCCCGAAATCCATCA-3'	(#25)
6	F: 5'-GATTCTCTGTGAGGCTTCAGATGT-3'	(#26)
	R: 5'-GCCAACAACAGCCAGCTGCACACA-3'	(#27)
7	F: 5'-TCCCCACTTAACCACTTTGC-3'	(#28)
	R: 5'-ACTTCATAGCTGACCATCTTCC-3'	(#29)

Primers for cDNA templates:

F:5'-GCCGCCCTCCTGCTGCTTCT-3'   (#30)

R: 5'-GGCTTAGGATTGGGGTCTGA-3'(#31)

We analyzed raw sequence data, generated consensus sequences, and produced nucleotide/amino acid alignments (DNASIS V2.6 for Windows, Hitachi Software Engineering Co., Ltd.). Mutations were detected by creating nucleotide/amino acid alignments of reported cathepsin C sequence data versus affected PLS patients sequence data using the Higgins-Sharpe UPGMA. Numbers in parentheses are SEQ ID NOS:.

Example I

PLS families

Parents of most families were consanguineous. Linkage studies localized a PLS gene in these five families to chromosome 11q14 [8]. Most affected individuals were homozygous for SSTR markers within the PLS candidate interval on chromosome 11q14, consistent with inheritance of both maternal and paternal copies of this genetic interval from a common familial ancestor ("identical by descent"). Four different haplotypes for short sequence tandem repeat (SSTR) markers spanning the

critical region were identified (Fig. 2), consistent
with four independent mutations in the gene responsible
for PLS.

5   **Analysis of cathepsin C**
              Using RT-PCR, we found cathepsin C is normally
       expressed in epithelium from palms, soles, knees and
       keratinized oral gingiva from unaffected individuals
       (data not shown). The cathepsin C gene spans
10   approximately 46 kb and consists of 7 exons. Sequence
       analysis of exonic, intronic and the 5'regulatory
       regions of the cathepsin C gene revealed PLS affected
       individuals from these families were homozygous for CTSC
       mutations that significantly altered the cathepsin C
15   open reading frame.
              Exon 6: Two affected individuals from one family
       were found to have an exon 6 nonsense mutation (856C->T)
       which introduces a stop codon at amino acid 286 (Fig.
       3).
20   *Exon 7*: Three different exon 7 mutations were
       detected (Fig. 4). A deletion of a single nucleotide
       (1047delA) of codon 349 was found that introduced a
       frame shift and an early termination codon (TGA) 27
       bases downstream. This mutation would result in a
25   mutated protein of 358 amino acids, compared to the
       normal (wild type) 463 amino acids. A deletion of 2
       bases of codon 343 (1028-1029delCT) resulting in the
       introduction of an early termination codon (TGA), and a
       truncated protein of 342 amino acids was identified in
30   another family. A G->A substitution in codon 429 (1286G-
       >A) that altered the original TRP codon (TGG) to a
       terminator codon (TAG) was identified in two affected
       individuals (#7 and #22) from two additional families.
       The expected truncated protein is 428 amino acids.
35   Although these families were not known to be related,
       the fact that affected individuals from these two
       Turkish families are homozygous for a common cathepsin C

gene mutation and also share a common haplotype for SSTR markers in the PLS candidate interval flanking the cathepsin C gene (D11S931 – D11S1311) suggests that these individuals have inherited the same cathepsin C

5   gene mutation from a common ancestor. Additional mutations were also identified in Exons 2, 3, 4, 5, 6 and 7. Summaries of the mutations identified to date are set forth in Table I and locations are shown in Figure 5.

10   Papillon Lefevre syndrome is a palmoplantar keratoderma (PPK) with the characteristic clinical features of palmoplantar hyperkeratosis, and severe periodontal destruction. The PPKs are a heterogeneous group of diseases all having gross thickening of the

15   palmoplantar skin. Clinically, the finding that distinguishes PLS from other PPKs is severe, early onset periodontal destruction. In affected individuals, the development and eruption of the primary teeth proceed normally, but the eruption of these teeth into the oral

20   cavity is associated with gingival inflammation and subsequent rapid destruction of the periodontium. This form of destructive periodontitis is characteristically unresponsive to traditional periodontal treatment modalities, and consequently, the primary dentition is

25   usually exfoliated prematurely. After exfoliation, the inflammation subsides, and the gingiva resumes a healthy appearance. However, with the eruption of the permanent dentition the process is usually repeated, resulting in the premature exfoliation of the permanent dentition,

30   although the third molars are sometimes spared [5]. Destruction of the alveolar bone in PLS is usually severe, resulting in generalized atrophy of the alveolar ridges, further complicating dental therapy.

         Because cathepsin C both localized to the refined

35   PLS candidate interval on chromosome 11q14 and was normally expressed in epithelium from sites affected by PLS it was evaluated as a candidate gene for PLS.

Cathepsin C, or dipeptidyl aminopeptidase I (EC 3.4.14.1), is a lysosomal cysteine protease capable of removing dipeptides from the terminus of protein substrates, but at higher pH it also exhibits dipeptidyl transferase activity [10]. The cathepsin C gene spans approximately 46 kb and consists of 7 exons that encode a 463-amino acid polypeptide with predicted features of the papain family of cysteine proteases [11]. Unlike cathepsin B, H, L, and S, which are small monomeric enzymes, cathepsin C is a large (200 kD) oligomeric protein that consists of four identical subunits, each composed of three different polypeptide chains [12,13]. Expression of cathepsin C (CTSC) is tissue dependent [14]. CTSC is expressed in pituitary gland, spinal cord, aorta, left atrium, right atrium, left ventricle, right ventricle, inter ventricular septum, apex of heart, esophagus, stomach, duodenum, jejunum, ileum, ileocecum, appendix, ascending colon, transverse colon, descending colon, rectum, kidney, skeletal muscle, spleen, thymus, peripheral blood lymphocytes, lymph node, bone marrow, trachea, lung, placenta, bladder, uterus, testis, liver, pancreas, adrenal gland, thyroid gland, salivary gland, mammary gland, fetal heart, fetal kidney, fetal liver, fetal spleen, fetal thymus, and fetal lung. The CTSC message is also expressed at high levels in immune cells including polymorphonuclear leukocytes and alveolar macrophages and is also expressed at high levels in osteoclasts [11,16]. The pathologic clinical findings of the PLS affected individuals studied here involve severe inflammation and destruction of the gingiva as well as hyperkeratosis of the skin from palmar, plantar and knee sites. In unaffected individuals, cathepsin C is normally expressed in epithelial tissues from sites clinically affected by PLS.

Most parents of the PLS affected individuals in this study are consanguineous. As a result, most PLS

affected individuals in each family are homozygous for
cathepsin C mutations inherited from a common ancestor.
Yet parents and several siblings who are heterozygous
carriers for cathepsin C mutations do not appear to show
5   either the palmoplantar hyperkeratosis or severe early
onset periodontitis characteristic of PLS. It appears
that the presence of one wild type cathepsin C gene is
sufficient to prevent PPK and periodontal destruction in
most patients. However, 1 mutation identified to date,
10   (1047 A->G) appears to be associated with the presence
of dermatological lesions. A consistent finding in the
three linkage reports to date is the lack of a common
haplotype among affected individuals from different
families. The present report describes 20 different
15   cathepsin C gene mutations associated with PLS. These
findings suggest that the CTSC mutations responsible for
PLS have arisen independently. Several mutations
reported here result in the introduction of premature
stop codons. While the W429X mutation encodes a protein
20   shortened by only 35 amino acids, the introduced stop
codon is 1 amino acid from the asparagine residue in the
active site (Fig. 5). It is likely that such a mutation
would cause a conformational alteration that may
decrease or abolish activity. Additionally, we have also
25   identified single nucleotide changes that result in
missense amino acid changes in several additional PLS
affected individuals from other populations, suggesting
that structural alterations of cathepsin C may cause
PLS.
30   In addition to the cardinal features of PLS,
reports suggest some PLS patients have an increased
susceptibility to infections [5]. This generalized
increased susceptibility to infection may reflect the
more deleterious effects of specific cathepsin C
35   mutations, or may reflect the epigenetic effects of
other gene loci. A variety of immunological findings
have been reported in PLS affected individuals including

decreased monocyte chemotaxis, decreased neutrophil chemotaxis, impaired neutrophil phagocytosis, altered superoxide production, and decreased blastogenic response, but it has been difficult to extrapolate results of these studies. Consequently, the underlying pathogenesis of PLS has been poorly understood [17]. Immunological findings previously reported for affected individuals from the current families includes decreased PMN chemotaxis and elevated CD11b expression [18,19]. The pathologic clinical findings associated with PLS suggest that cathepsin C is functionally important in the structural growth and development of skin and in susceptibility to periodontal disease. As a lysosomal cysteine proteinase, cathepsin C is important in intracellular degradation of proteins and appears to be a central coordinator for activation of many serine proteinases in immune/inflammatory cells [11]. It is unknown if the profound periodontal disease susceptibility is a consequence of altered integrity of junctional epithelium surrounding the teeth. It is interesting that once teeth are exfoliated, and consequently the junctional epithelium is eliminated, the severe gingival inflammation resolves. A more complete understanding of the functional physiology of cathepsin C carries significant implications for understanding periodontal disease susceptibility. Identification of cathepsin C gene mutations in PLS raises the possibility of creating an animal model to study the development, treatment and prevention of hyperkeratosis and periodontitis.

Classification of the PPKs based upon histological findings, epidermolysis and localization of lesions within the skin (diffuse, linear or focal) has not been helpful in understanding the pathomechanism of disease [20]. Identification of mutations in specific genes has led to development of a revised nosology of these diseases in which PLS is grouped with the palmoplantar

ectodermal dysplasias [2]. In addition to providing
insight into both normal as well abnormal epithelial
growth and development, identification of mutations in
cathepsin C associated with PLS will contribute to the

5   overall nosology of the PPKs.

EXAMPLE II

CTSC Mutation in Haim-Munk Syndrome

Of the many palmoplantar keratoderma (PPK)

10 conditions, only Papillon Lefevre syndrome (PLS) and
Haim Munk syndrome (HMS) are associated with premature
periodontal destruction. Although both PLS and HMS share
the cardinal features of PPK and severe periodontitis, a
number of additional findings are reported in HMS

15 including arachnodactyly, acroosteolysis, atrophic
changes of the nails, and a radiographic deformity of
the fingers. While PLS cases have been identified
throughout the world, HMS has only been described among
descendents of a religious isolate originally from

20 Cochin, India. Parental consanguinity is a
characteristic of many cases of both conditions.
Although autosomal recessive transmission of PLS is
evident, a more "complex" autosomal recessive pattern of
inheritance with phenotypic influences from a closely

25 linked modifying locus has been hypothesized for HMS. As
set forth in Example I, mutations of the cathepsin C
gene have been identified as the underlying genetic
defect in PLS. To determine if a cathepsin C mutation is
also responsible for HMS, we sequenced the gene in

30 affected and unaffected individuals from families with
HMS. Here we report identification a mutation of
cathepsin C (exon 6, 857A->G) that changes a highly
conserved amino acid in the cathpesin C peptide. This
mutation segregates with HMS in four nuclear families.

35 Additionally, the existence of a shared common haplotype
for genetic loci flanking the cathepsin C gene suggests
that affected individuals descended from the Cochin

isolate are homozygous for a mutation inherited "identical by descent" from a common ancestor. This finding supports simple autosomal recessive inheritance for HMS in these families. As described above, we also

5   report a mutation of the same exon 6 CTSC codon (856C->T) in a Turkish family with classic PLS. These findings provide evidence that PLS and HMS are allelic variants of cathepsin C gene mutations.

In addition to congenital palmoplantar keratosis

10   and progressive early onset periodontal destruction, other clinical findings shared by these individuals included recurrent pyogenic skin infections, acroosteolysis, atrophic changes of the nails, arachnodactyly, and a peculiar radiographic deformity of

15   the fingers consisting of tapered pointed phalangeal ends and a clawlike volar curve (Figures 6A and 6B). Subsequently pes planus was reported to be associated with the syndrome [24]. This was the first reported association of these clinical findings, and the

20   condition became known as Haim Munk syndrome, or keratosis palmoplantaris with periodontopathia and onychogryposis (HMS1; MIM245010)[22]. Although the palmoplantar findings and severe periodontitis were suggestive of the Papillon-Lefevre syndrome (PLS;

25   MIM245000)[3], the association of other clinical features, particularly nail deformities and arachnodactyly, argued that HMS was a distinct disorder. In contrast to PLS, the skin manifestations in HMS were reported to be more severe and extensive. In addition to

30   a marked palmoplantar keratosis (Figure 6C, 6D), affected individuals had scaly erythematous and circumscribed patches on the elbows, knees, forearms, shins and dorsum of the hands. While the periodontium in HMS was reported to be less severely affected than in

35   PLS, gingival inflammation and alveolar bone destruction are present and severe (Figure 6E, 6F). In a subsequent genetic study of this extended family, Hacham-Zadeh and

62

coworkers [25] concluded that the syndrome might not behave as a simple autosomal recessive trait. Based upon their estimate of the disease allele frequency in this population (0.1), the absence of the condition in other kindreds of the Cochin isolate, and an inability to document consanguinity for many of the parents of affected individuals, they hypothesized that a "complex" autosomal recessive inheritance pattern with a closely linked dominant modifier locus may be responsible for the HMS phenotype.

HMS families

Pedigrees of the reported familial relationships for the Cochin descendents are shown in Figure 7A. Descendents of the Cochin isolate studied include sibships 2, 3, 4 and 5 in the kindred pedigree originally described by Hacham-Zadeh and coworkers [25].

HMS family genotyping results

All HMS affected individuals from the Cochin kindred were found to be homozygous for all three polymorphic DNA loci (D11S1887, D11S1780 and D11S1367) flanking the cathepsin C locus. Additionally, these individuals shared a common haplotype for these polymorphic markers. These findings are consistent with inheritance of both maternal and paternal copies of this genetic interval from a common familial ancestor ("identical by descent").

Analysis of cathepsin C in HMS

The cathepsin C gene spans approximately 46 kb and consists of 7 exons. Sequence analysis of exonic, intronic and the 5'regulatory regions of the cathepsin C gene revealed that HMS affected individuals from the Cochin kindred were homozygous for a mutation in codon 286 of exon 6 (857A->G) which results in substitution of a conserved glutamine residue at position 286 by an

arginine: Q286R (Figure 8). This glutamine residue is normally completely conserved in wild type cathepsin C from at least five species (data not shown). This was the only sequence change different from the reported,

5 highly conserved, wild type CTSC sequence (GenBank Accession No.: AC011088; SEQ ID NO: 1). All available parents of HMS affected individuals were found to be heterozygous for the mutated (857A->G) allele and the wild type allele. None of the parents or siblings

10 heterozygous for the mutated (857A->G) allele and the wild type allele manifested clinically identifiable characteristics of PPK or had a history of severe, early onset periodontitis.

15 **Restriction Analysis**

The Q286R mutation creates an AvaI restriction cleavage site. We utilized this newly created restriction site to develop a rapid test to screen for the Q286R mutation. After amplification of a 465bp

20 fragment encompassing the 3' end of exon 6 using primers: Forward 5'-GTATGCTAGAAGCGAGAATCCGTAT-3' (SEQ ID NO: 32) and Reverse 5'-CCAATGCTAAAACTTGTTGAGACC-3' (SEQ ID NO: 33), the PCR products were purified using the Promega PCR kit according to the manufacturer's

25 instructions. Purified products were eluted in 20 µl water. Approximately 5-10 µl of purified product was digested with 5U AvaI (New England Biolabs) in a total volume of 15 µl for 1.5 hr at 37°C. Following digestion, the products were separated by

30 electrophoresis through an 1.8 % agarose gel. Amplification of the wildtype sequence results in a 465 bp product that is not cleaved by AvaI. Amplification of the mutated (857A(G) sequence results in a 465 bp product that is cleaved by AvaI to yield products of 404

35 and 61 bp. Accordingly, individuals who are homozygous for the wildtype sequence exhibit a 465 bp band. Heterozygous individuals exhibit 3 bands: 465, 404, and

61 bp bands. Individuals who are homozygous for the
Q286R mutation exhibit bands of 404 and 61 bp.
Restriction analysis confirmed the sequencing results of
all examined individuals (Figure 9).

5

EXAMPLE III

Genetic Screening for PPK-Associated Mutations

The foregoing findings provide the basis for
screening and diagnostic assays for assessing patients
10   for the presence of mutations in the CTSC gene related
to the pathological conditions described herein. A
summary of the mutations in CTSC identified as
associated with PPKs are set forth in Table 1.

15

Table 1. Phenotype correlations with *CTSC* mutations

Mutation[a]	Predicted Effect	Location	Phenotype[b]	# of Families
199-222del	Deletion of aa 67-74	Exon 2	PLS	1
445-446insATGT	Frameshift, termination at aa 157	Exon 3	PLS	1
458C®T	Threonine 153 Isoleucine	Exon 3	PLS	2
622-623insC	Frameshift, termination at aa 223	Exon 4	PLS	1
704G®A	Tryptophan 235 Stop	Exon 5	PLS	1
748C®T	Arginine 250 Stop	Exon 5	PLS	1
815G®C	Arginine 272 Proline	Exon 6	PLS	1
856C®T	Glutamine 286 Stop	Exon 6	PLS	2
857A®G	Glutamine 286 Arginine	Exon 6	HMS	1
898G®A	Glycine 300 Serine	Exon 7	PLS	1
901G®T	Glycine 301 Valine	Exon 7	PLS	1
901G®A	Glycine 301 Serine	Exon 7	PLS	1
910T®A	Tyrosine 304 Asparagine	Exon 7	PLS, RP	1
956A®G	Glutamic Acid 319 Glycine	Exon 7	PLS	1
1015C®T	Arginine 339 Cysteine	Exon 7	PLS	2
1019A®G	Tyrosine 340 Cysteine	Exon 7	PLS	1
1028-1029delCT	Introduction of premature termination codon	Exon 7	PLS	1
1040A®G	Tyrosine 347 Cysteine	Exon 7	PPP	2
1047delA	Frameshift, termination at aa 349	Exon 7	PLS	1
1286G®A	Tryptophan 429 Stop	Exon 7	PLS	3
1360A®G	Glutamic Acid 447 Glycine	Exon 7	PLS	1

1. cDNA numbering considering the initiator Met codon as nucleotide +1.
2. Phenotype symbols: PLS, Papillon-Lefevre syndrome; PPP, Prepubertal periodontitis;
   HMS, Haim Munk syndrome; RP, Retinitis pigmentosa

While the mutations described in the previous examples are associated with certain pathological conditions, it is important to note that the CTSC gene contains many polymorphisms. Many of these genetic changes are not associated with the disease state. The genetic changes assessed by the methods of the present invention must be associated with the production of an aberrant CTSC protein. Accordingly, a suitable assay for diagnosing this disorder includes the step of differentiating harmless polymorphisms from those

mutations which give rise to PPKs and periodontal
disorders. These include changes in the coding sequence
which give rise to decreased mRNA stability as compared
to wild type CTSC mRNA. Alternatively cathepsin C

5   enzymatic activity can be compared between altered CTSC
coding sequences and nucleic acids encoding the wild
type enzyme. Such assays are well known in the art and
need not be set forth here. See for example, McGuire et
al., Archives of Biochemistry and Biophysics 295:280-8,

10 1992; McDonald et al., J. of Biological Chemistry
244:2693-26709, 1969; Metroione et al, Biochemistry
5:1597-1604, 1966; and Vanha-Perttula et al.,
Histochemie 5:170-181, 1965.

15   **References:**

1.   Papillon, M.M., and Lefevre, P. Deux cas de
keratodermie palmaire et plantaire symetrique familiale
(maladie de Meleda) chez le frere et la soeur.

20 Coexistence dans les deus cas d'alterations dentaires
grabes. *Bull Soc Fr Dermatol Syphilis* 1924;**31**:82-87.

2. Stevens, H.P., Kelsell, D.P., Bryant, S.P., Bishop,
D.T., Spurr, N.K., Weissenbach, J., Marger, D., Marger,

25 R.S., Leigh, I.M. Linkage of an American pedigree with
palmoplantar keratoderma and malignancy (palmoplantar
ectodermal dysplasia type III) to 17q24. Literature
survey and proposed updated classification of the
keratodermas. *Arch Dermatol*. 1996;**132**:(6) 640-51.

30 3. Online Mendelian Inheritance in Man, OMIM (TM).
Johns Hopkins University, Baltimore, MD. MIM Number:
{MIM 245000}: {8/20/98}: *World Wide Web URL*:
http://www.ncbi.nlm.nih.gov/omim/

35 4.   Gorlin, RJ, Sedano H, and Anderson VE (1964),
RJ. The syndrome of palmar-plantar hyperkeratosis and
premature periodontal destruction of the teeth. *J
Pediatr* 1964; **65**:895-908.

40 5. Haneke, E. The Papillon-Lefevre syndrome:
keratosis palmoplantaris with periodontopathy. Report of
a case and review of the cases in the literature. *Hum
Genet* 1979;**51**:(1) 1-35.

45 6. Laass, M.W., Hennies, H.C., Preis, S., Stevens,
H.P., Jung, M., Leigh, I.M., Wienker, T.F., Reis, A..

Localization of a gene for Papillon-Lefevre syndrome to chromosome 11q14-q21 by homozygosity mapping. *Hum Genet* 1997;**101**:(3) 376-82.

7. Fischer, J., Blanchet-Bardon, C., Prud'homme, J.F., Pavek, S., Steijlen, P.M., Dubertret, L., Weissenbach, J. Mapping of Papillon-Lefevre syndrome to the chromosome 11q14 region. *Eur J Hum Genet 1997;* **5**:(3)156-60.

8. Hart, T.C., Bowden, D.W., Ghaffar, K.A., Wang, W., Cutler, C.W., Cebeci, I., Efeoglu, A., Firatli, E. Sublocalization of the Papillon-Lefevre syndrome locus on 11q14-q21. *Amer J Med Genet* 1998;**79**:(2) 134-139.

9. Hart TC, Bowden DW, Hart PS, Walker SJ, Callison SA, Bobby PL, Firatli E. An integrated physical and genetic map of the PLS locus interval on chromosome 11q14. Mammalian Genome *In Press*

10. Kirschke, H., Barrett, A.J., Rawlings, N.D. Proteinases 1: Lysosomal cysteine proteinases. In: Protein Profile **Vol.2**, P. Sheterline, ed. (London, UK: *Academic Press Ltd.)* 1995;1587-1643.

11. Rao, N.V., Rao, G.V., Hoidal, J.R. Human dipeptidyl-peptidase I. *J Biol Chem*.1997; **272**:10260-10265.

12. Paris, A., Strukelj, B., Pungercar, J., Renko, M., Dolenc, I., Turk, V. Molecular cloning and sequence analysis of human preprocathepsin C. *FEBS Lett* 1995;**369**:326-330.

13. Dolenc, I., Turk, B., Pungercic, G., Ritonja, A., Turk, V. Oligomeric structure and substrate induced inhibition of human cathepsin C. *J Biol Chem* 1995;**270**:21626-31.

14. Pham, C. T. N., Armstrong, R. J., Zimonjic, D.B., Popescu, N. C., Payan, D. G., Ley, T. J. Molecular cloning, chromosomal localization, and expression of murine dipeptidyl peptidase I. *J Biol Chem* 1997;**272**:10695-10703.

15. McGuire, M.J., Lipsky, P.E., Thiele, D.L. Cloning nad characterization of the cDNA encoding mouse dipeptidyl peptidase I (cathepsin C). *Biochimica et Biophysica Acta* 1997; **1351**:267-73.

16. Hakeda, Y., Kumegawa, M. Osteoclasts in bone metabolism. Kaibogaku Zasshi *J Anat* 1991;**66**:215-225.

17. Hart , T.C. and Shapira, L. Papillon Lefevre syndrome. *Periodontol 2000* 1994;**6**, 88-100.

18. Firatli, E., Tuzun, B., Efeoglu, A. Papillon-Lefevre syndrome. Analysis of neutrophil chemotaxis. *J Periodontol* 1996;**67**:617-20.

19. Firatli, E., Gurel, N., Efeoglu, A., Badur, S. Clinical and immunological findings in 2 siblings with Papillon-Lefevre syndrome. *J Periodontol* 1996;**67**:1210-5.

20. Itin, P.H. Classification of autosomal dominant palmoplantar keratoderme: past-present-future. *Dermatol* 1992;**185**:163-165.

21. Haim,S.,Munk J: Keratosis palmo-plantaris congenita, with periodontosis, arachnodactyly and peculiar deformity of the terminal phalanges. *Br J Dermatol* 1965;**77**:42-45.

22. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {MIM 245010}: {5/20/97}: *World Wide Web URL*: http://www.ncbi.nlm.nih.gov/omim/

23.Gorlin, R. J., Pindborg, J. J., Cohen, M. M., Jr. Syndromes of the Head and Neck. New York: McGraw-Hill (pub.) (2nd ed.) 1976. Pp. 373-376.

24. Puliyel JM, Sridharan Iyer KS. A syndrome of keratosis palmo-plantaris congenita, pes planus, onychogryphosis, periodontosis, arachnodactyly and a peculiar acro-osteolysis. Brit J Derm 1986;115:243-248.

25. Hacham-Zadeh S, Schaap T, Cohen MM. A genetic analysis of the Papillon-Lefevre syndrome in a Jewish family from Cochin. Am J Med Genet 1978;2:153-157.

26 Gorlin RJ,Cohen MM jr., Levin LS. Syndromes of the head and neck . 3rd edn. New York. Oxford University Press, 1990:853-854 (Oxford Monographs on Medical Genetics, No.19).

27. Hart T, Hart PS, Bowden D, Walker S, Callison S, Michalec M, Firatli E. Mutations of the cathepsin C gene are responsible for Papillon Lefevre syndrome. In Press J Med Genet

28. Hart TC, Stabholz A, Meyle J, Shapira L, Van Dyke TE, Cutler CW, Soskolne WA. Genetic studies of syndromes with severe periodontitis and palmoplantar hyperkeratosis. J Periodont Res 1997;32:81-89.

29. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M. A second-generation linkage map of the human genome.

69

Nature 1992;359:794-801.

30.   Lucker GP, Van de Kerkhof PC, Steijlen PM. The hereditary palmoplantar keratoses: an updated review and classification. Brit J Derm 1994;131(1):1-14.

31. Smith P, Rosenzweig KA. Seven cases of Papillon Lefevre Syndrome. Periodontics 1967;5:42-6.

32.   Cohen T, Bloch N. Immigrant Jews from Cochin. In Goldschmidt E (ed) The Genetics of Migrant and Isolate Populations. Baltimore: Williams & Wilkins 1963;352.

33. Wilkie AO. Craniosynostosis: genes and mechanisms. Hum Molec Genet 1997;6:1647-56.

34. Hola-Jamriska L, Tort JF, Dalton JP, Day SR, Fan J, Asskov J and Brindley PJ. Cathepsin C from Schistosoma japonicum; cDNA encoding the preproenzyme and its phylogenetic relationships. Eur J Biochem 1998;255:527-534.

35.   Wolters PJ, Raymond WW, Blount JL, Caughey GH. Regulated expression, processing, and secretion of dog mast cell dipeptidyl peptidase I. J Biol Chem 1998;25:15514-15520.

36. Mackenzie IC, Rittman G, Gao Z, Leigh I, Lane EB. Patterns of cytokeratin expression in human gingival epithelia. J Periodont Res 1991;26468-478.

37. Hormia M, Sahlberg C, Thesleff I, Airenne T. The epithelium-tooth interface--a basal lamina rich in laminin-5 and lacking other known laminin
     isoforms. J Den Res (1998);77:1479-85.

38. Stabholz A, Taichman NS, Soskolne WA. Occurrence of Actinobacillus actinomycetemcomitans and anti leukotoxin antibodies in some members of an extended family affected by Papillon-Lefevre syndrome. J Periodontol 1995;66:653-57.

39.   Page RC. Altman LC. Ebersole JL. Vandesteen GE. Dahlberg WH. Williams BL. Osterberg SK. Rapidly progressive periodontitis. A distinct clinical condition. Journal of Periodontology. 54(4):197-209, 1983

40. Page RC. Bowen T. Altman L. Vandesteen E. Ochs H. Mackenzie P. Osterberg S. Engel LD. Williams BL. Prepubertal periodontitis. I. Definition of clinical disease entity. Journal of Periodontology. 54(5):257-71, 1983

70

While certain of the preferred embodiments of the
present invention have been described and specifically
exemplified above, it is not intended that the invention
be limited to such embodiments. Various modifications

5   may be made thereto without departing from the scope and
spirit of the present invention, as set forth in the
following claims.

10

What is claimed is:

1.   An nucleic acid molecule encoding an altered CTSC protein, said nucleic acid having at least one of the alterations set forth in Table 1.

2.   A nucleic acid probe specifically hybridizable to a human altered CTSC-encoding nucleic acid and not to wild-type CTSC encoding nucleic acids, said altered CTSC encoding nucleic acid having one of the alterations set forth in Table 1.

3. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

4. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

5. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

6. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby causing a premature stop codon.

7. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a

72

substitution of a G for a A at nucleotide position 1286
in Exon 7, thereby replacing a tryptophan codon with a
premature stop codon.

5   8.   The nucleic acid probe of claim 2 wherein said
altered CTSC DNA has the alteration comprising a
substitution of a C for a T at nucleotide position 1015
in Exon 7, thereby replacing a codon encoding arginine
for a cysteine encoding codon.

10

          9.   The nucleic acid probe of claim 2 wherein said
altered CTSC DNA has the alteration comprising a
substitution of an A for a G at nucleotide position 1019
in Exon 7, thereby replacing a codon encoding tyrosine
15   for a cysteine encoding codon.

          10.   The nucleic acid probe of claim 2 wherein
said altered CTSC DNA has the alteration comprising a
substitution of an A for a G at nucleotide position 1040
20   in Exon 7, thereby replacing a codon encoding tyrosine
for a cysteine encoding codon.

          11.   A mutated CTSC protein encoded by a CTSC
encoding nucleic acid, said nucleic acid containing a
25   mutation as set forth in Table 1.

          12.   An antibody immunologically specific for the
protein of claim 11.

30   13. A method for detecting a germline alteration in
a CTSC gene, said alteration selected from the group
consisting of the alterations set forth in Table 1 in a
human, said method comprising analyzing a sequence of a
CTSC gene or CTSC RNA from a human sample or analyzing a
35   sequence of CTSC cDNA made from mRNA from said human sample.

          14. The method of claim 13 which comprises
                              73

analyzing CTSC RNA from the subject.

15. The method of claim 14 wherein a germline alteration is detected by hybridizing a CTSC gene probe which specifically hybridizes to nucleic acids containing at least one of said alterations and not to wild-type CTSC sequences to RNA isolated from said human sample and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration in said RNA and thereby the presence of said germline alteration in said sample.

16. The method of claim 13 wherein a germline alteration is detected by obtaining a first CTSC gene fragment from a CTSC gene isolated from said human sample and a second CTSC gene fragment from a wild-type CTSC gene, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment, electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first CTSC gene fragment is shifted relative to said second CTSC gene fragment and sequencing said single-stranded DNA from said first CTSC gene fragment having a shift in mobility.

17. The method of claim 13 wherein a germline alteration is detected by hybridizing a CTSC probe which specifically hybridizes to nucleic acids containing at least one of said alterations and not to wild-type CTSC sequences to genomic DNA isolated from said sample and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said germline alteration in the sample.

18. The method of claim 13 wherein a germline

74

alteration is detected by amplifying all or part of a
CTSC gene in said sample using a set of primers specific
for a wild-type CTSC gene to produce amplified CTSC
nucleic acids and sequencing the amplified CTSC nucleic acids.

5

19. The method of claim 13 wherein a germline
alteration is detected by amplifying all or part of a
CTSC gene in said sample using a primer specific for an
allele having for one of said alterations and detecting

10   the presence of an amplified product, wherein the
presence of said product indicates the presence of said
allele in the sample.

20. The method of claim 13 wherein a germline
15   alteration is detected by molecularly cloning all or
part of a CTSC gene in said sample to produce a cloned
nucleic acid and sequencing the cloned nucleic acid.

21. The method of claim 13 wherein a germline
20   alteration is detected by forming a heteroduplex
consisting of a first strand of nucleic acid selected
from the group consisting of CTSC gene genomic DNA
fragment isolated from said sample, CTSC RNA fragment
isolated from said sample and CTSC cDNA fragment made
25   from mRNA from said sample and a second strand of a
nucleic acid consisting of a corresponding human
wild-type CTSC gene fragment, analyzing for the presence
of a mismatch in said heteroduplex, and sequencing said
first strand of nucleic acid having a mismatch.

30

22. The method of claim 13 wherein a germline
alteration is detected by amplifying CTSC gene nucleic
acids in said sample, hybridizing the amplified nucleic
acids to a CTSC DNA probe which specifically hybridizes
35   to nucleic acids containing at least one of said
alterations and not to wild-type CTSC sequences and
detecting the presence of a hybridization product,

wherein a presence of said product indicates the
presence of said germline alteration.

23. The method of claim 13 wherein a germline
alteration is detected by analyzing the sequence of a
CTSC gene in said sample for one of the mutations set
forth in Table 1.

24. The method of claim 13 wherein a germline
alteration is detected by obtaining a first CTSC gene
fragment from a CTSC gene isolated from said human
sample and a second CTSC gene fragment from a CTSC
allele specific for one of said alterations, said second
fragment corresponding to said first fragment, forming
single-stranded DNA from said first CTSC gene fragment
and from said second CTSC gene fragment,
electrophoresing said single-stranded DNAs on a
non-denaturing polyacrylamide gel and comparing the
mobility of said single-stranded DNAs on said gel to
determine if said single-stranded DNA from said first
CTSC gene fragment is shifted relative to said second
CTSC gene fragment, wherein no shift in electrophoretic
mobility indicates the presence of said alteration in
said sample.

26. The method of claim 13 wherein a germline
alteration is detected by obtaining a first CTSC gene
fragment from (a) CTSC gene genomic DNA isolated from
said sample, (b) CTSC RNA isolated from said sample or
(c) CTSC cDNA made from mRNA isolated from said sample
and a second CTSC gene fragment from a CTSC allele
specific for one of said alterations, said second
fragment corresponding to said first fragment, forming
single-stranded DNA from said first CTSC gene fragment
and from said second CTSC gene fragment, forming a
heteroduplex consisting of single-stranded DNA from said
first CTSC gene fragment and single-stranded DNA from

said second CTSC gene fragment and analyzing for the presence of a mismatch in said heteroduplex, wherein no mismatch indicates the presence of said alteration.

27. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

28. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

29. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

30. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby causing a premature stop codon.

31. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a G for a A at nucleotide position 1286 in Exon 7, thereby replacing a tryptophan codon with a premature stop codon.

32. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 1015 in Exon 7, thereby replacing a codon encoding arginine for a cysteine encoding codon.

33.   A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1019 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

34.   A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1040 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

35. A method for detecting a germline alteration in a CTSC human encoding nucleic acid, said method comprising comparing a sequence of a CTSC DNA or CTSC RNA from a human sample with an isolated wild type CTSC sequence as provided in SEQ ID NO:1.

36. A method as claimed in claim 35, wherein stability of said altered CTSC mRNA is compared with stability of wild type CTSC mRNA.

37.   A method as claimed in claim 35, further comprising expressing an altered CTSC protein from said altered CTSC encoding nucleic acid and comparing cathepsin C enzymatic activity of said altered CTSC protein to enzymatic activity of wild-type cathepsin C.

38.   A kit for detecting the presence of an altered CTSC encoding nucleic acid in a biological sample, comprising:

   i)   oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1;

   ii)   reaction buffer; and

   iii)   an instruction sheet.

39.   A kit as claimed in claim 38, wherein said oligonucleotide contains a tag.

40.   A kit for detecting the presence an altered CTSC encoding nucleic acid in a biological sample, comprising:

   i)   antibodies immunologically specific for the altered CTSC proteins of the invention;

   ii)   a solid support with immobilized CTSC antigens as a positive control; and

   iii)   an instruction sheet.

41.   A kit as claimed in claim 40, wherein said antibody contains a tag.

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 February 2001 (01.02.2001)

PCT

(10) International Publication Number
WO 01/07663 A1

(51) International Patent Classification[7]: **C12Q 1/68**, C12P 19/34, C07H 21/04, C07K 14/435, 16/18, G01N 33/53

(21) International Application Number: PCT/US00/20400

(22) International Filing Date: 27 July 2000 (27.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/145,644   27 July 1999 (27.07.1999)   US
60/165,016   12 November 1999 (12.11.1999)   US

(71) Applicant *(for all designated States except US)*: **WAKE FOREST UNIVERSITY** [US/US]; Medical-Center Boulevard, Winston-Salem, NC 27157 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **HART, Thomas, C.** [US/US]; 2476 Matterhorn Drive, Wexford, PA 15090-7612 (US).

(74) Agents: **RIGAUT, Kathleen, D.** et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103-2307 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES

(57) Abstract: Compositions related to mutations in the CTSC gene associated with palmoplantar keratodermas are disclosed. Also provided are methods for using the compositions of the invention in diagnostic and prognostic screening assays.
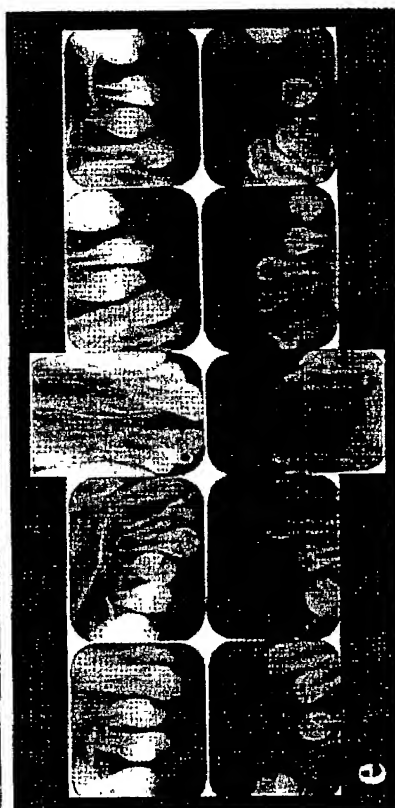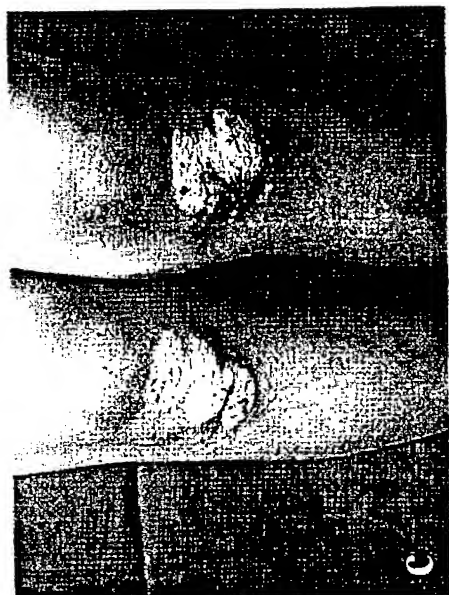
Figure 1

Figure 2

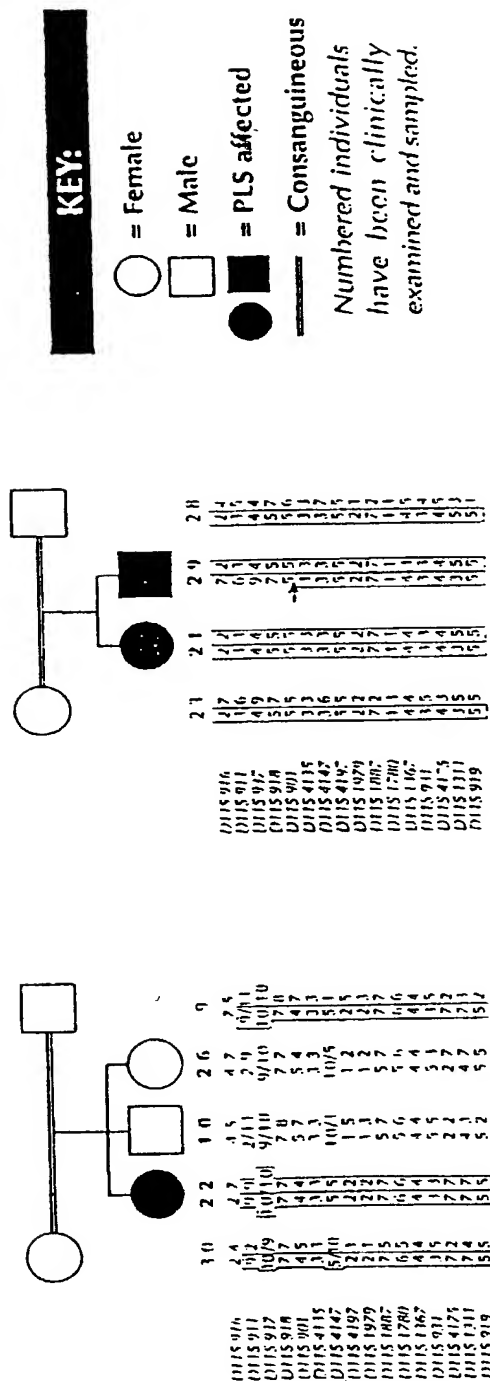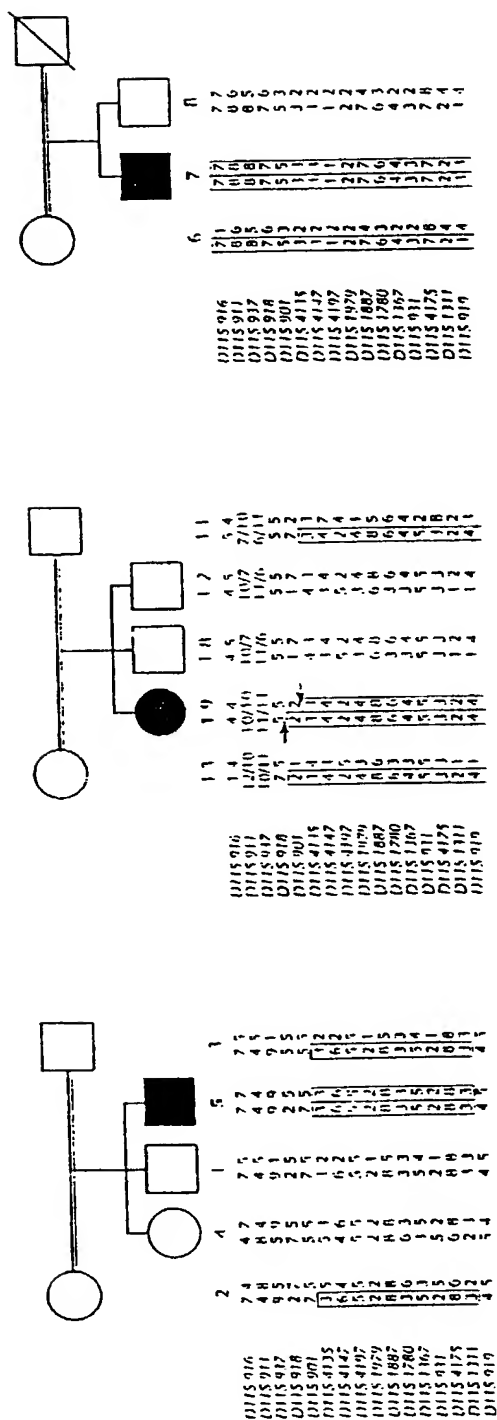Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

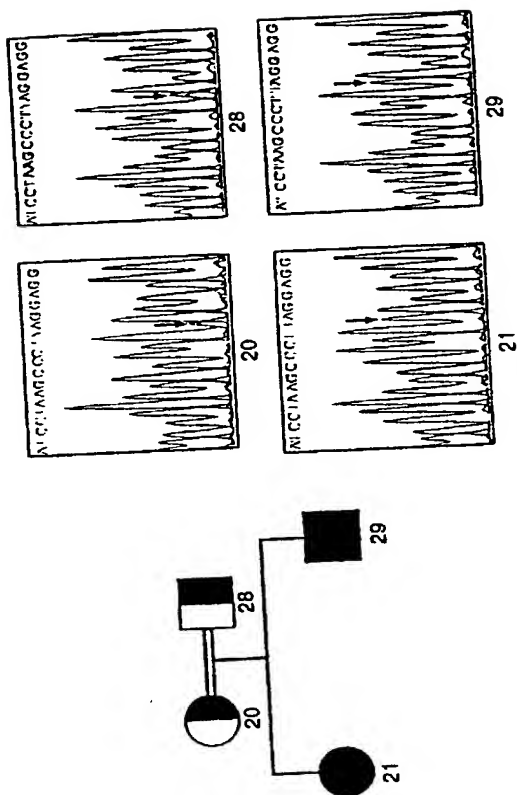Wildtype sequence from nt 2114 to 2133- ATCCTAAGCCCTCAGGAGG
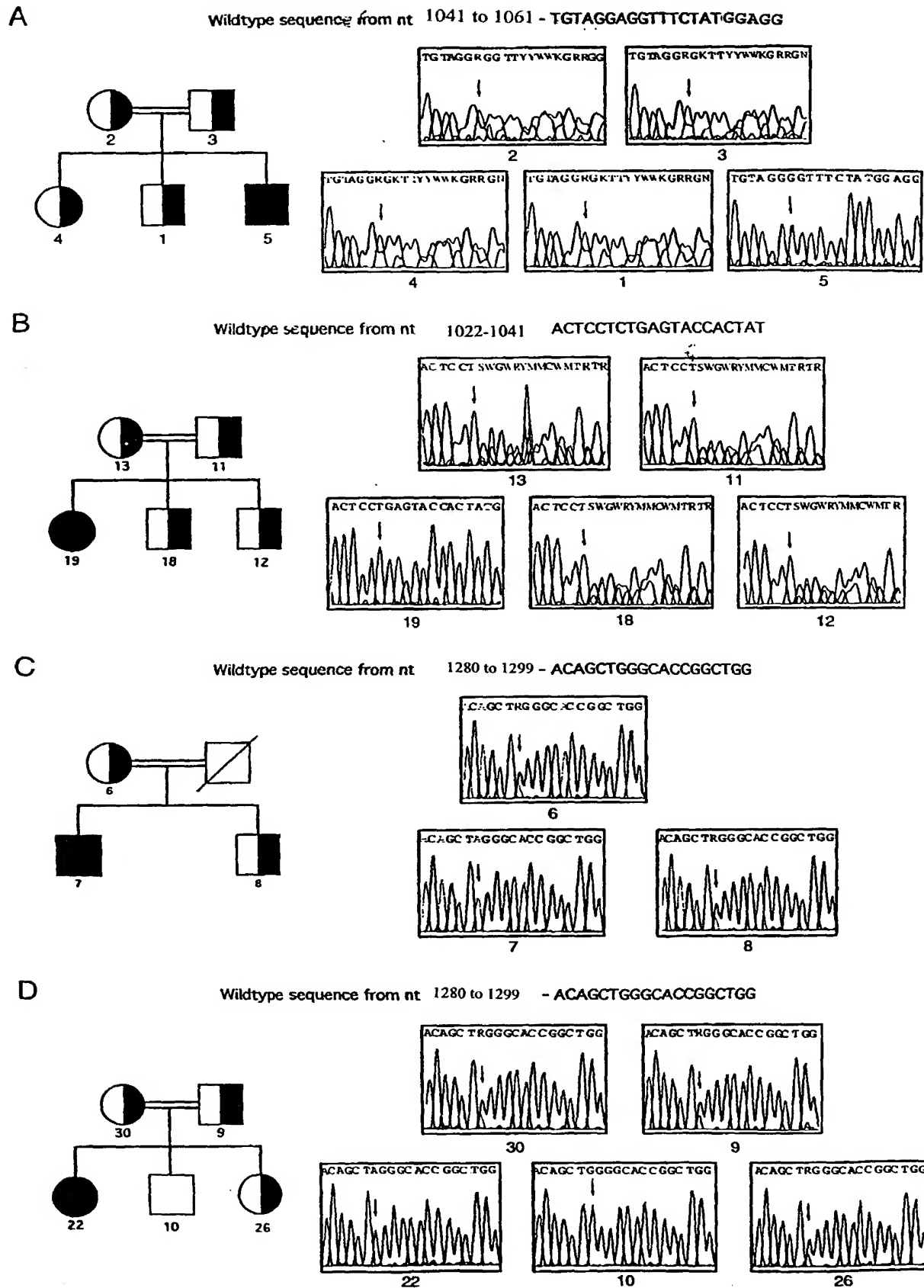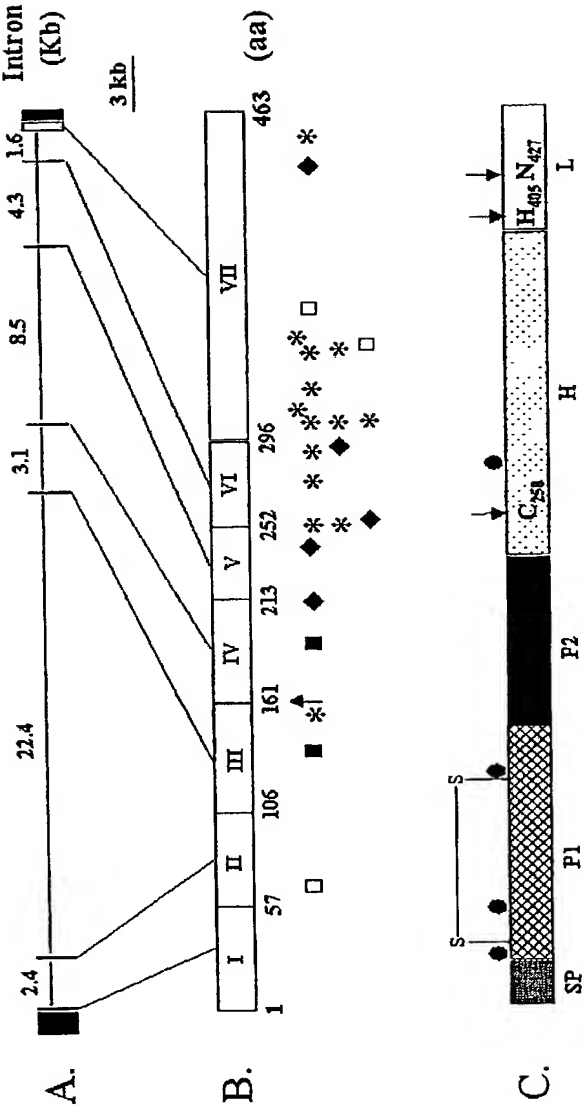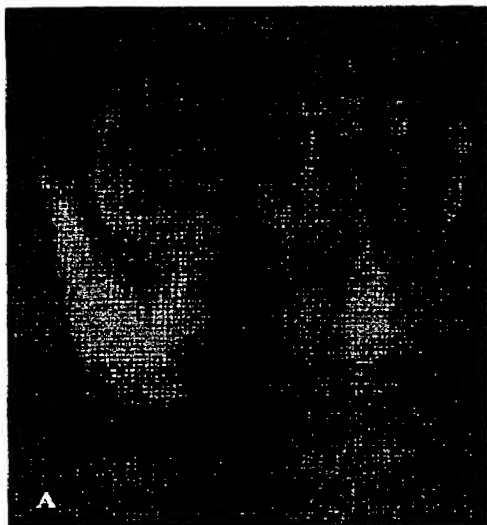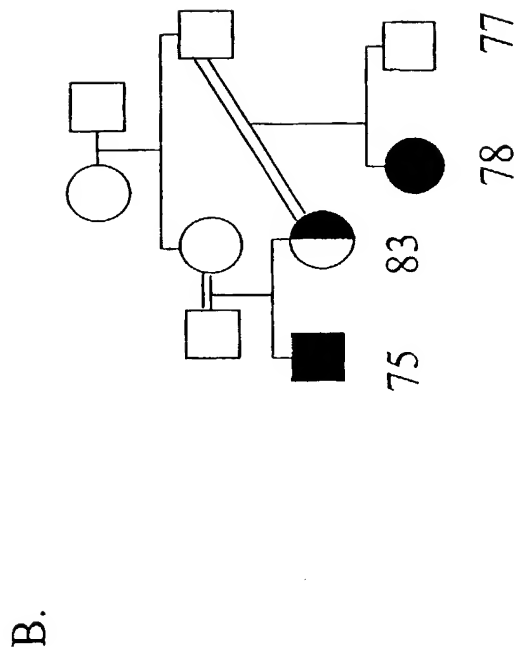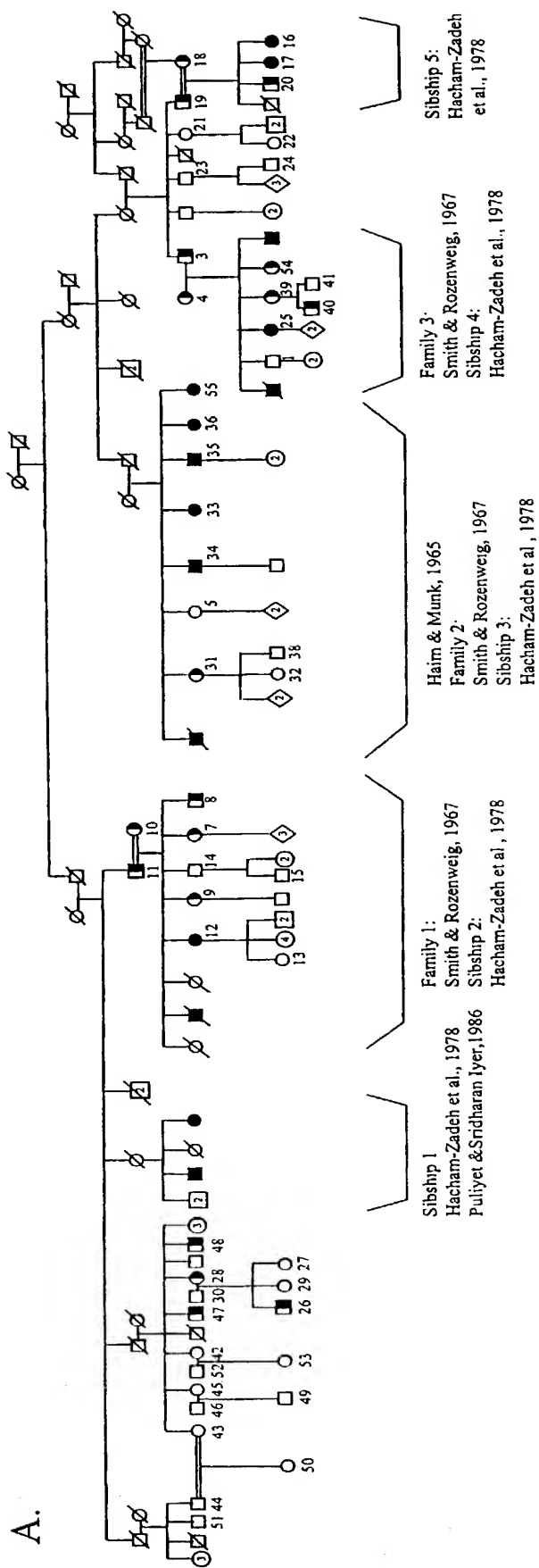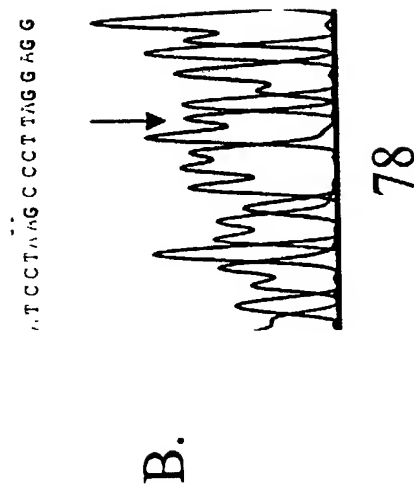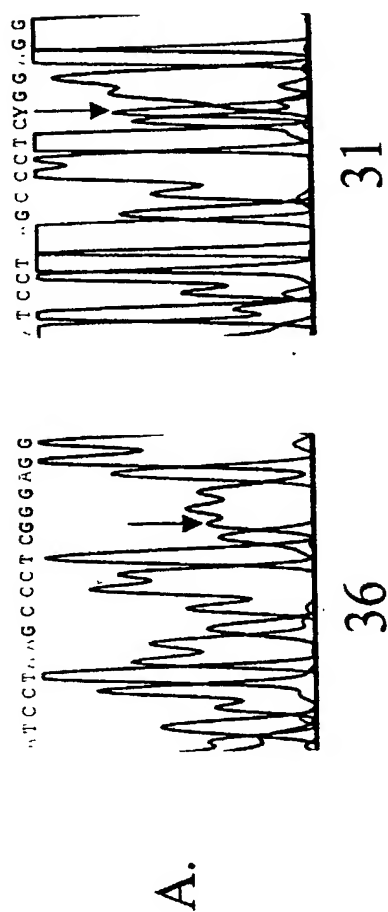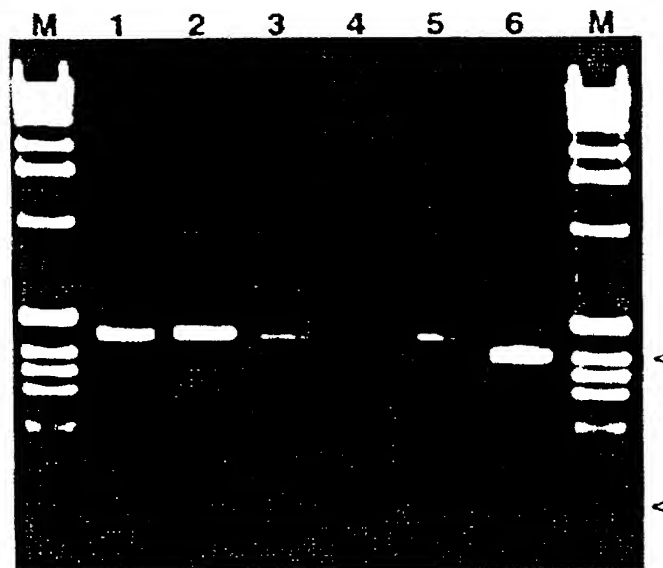
Figure 8

Figure 9

UTILITY
Original U.S. or PCT D/O

DECLARATION, POWER OF ATTORNEY AND POWER TO INSPECT

As a below named inventor, I hereby declare:

that my residence, post office address and citizenship are as stated below next to my name;

that I verify believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the invention entitled: METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES

the specification of which [check one(s) applicable]

  X   was filed July 27, 2000 as International Application No. PCT/US00/20400 on which U.S. Patent Application No. _____, filed January 25, 2002 is based.
_____ and was amended by Amendment filed _____ (if applicable); [or];
_____ is attached to this Declaration, Power of Attorney and Power to Inspect;

that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above; and that I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Rule 56 (a) [37 C.F.R. §1.56(a)]

CLAIM UNDER 35 USC §119(e): I hereby claim the benefit under 35 USC §119(e) of any United States provisional applications listed below:

Provisional Application No.	Filing Date Day/Mo/Year
60/144,644	27 July 1999
60/165,018	12 November 1999

POWER OF ATTORNEY: As inventor, I hereby appoint DANN, DORFMAN, HERRELL AND SKILLMAN, P.C. of Philadelphia, PA, and the following individual(s) as my attorneys or agents with full power of substitution to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: Kathleen D. Rigaut, Ph.D., J.D. Reg. No. 43,047; Maria Kourtakis, Esq. Reg. No. 41,126 and Patrick J. Hagan, Esq. Req. No. 37,643

POWER TO INSPECT: I hereby give DANN, DORFMAN, HERRELL AND SKILLMAN, P.C. of Philadelphia, PA or its duly accredited representatives power to inspect and obtain copies of the papers on file relating to this application.

SEND CORRESPONDENCE TO: CUSTOMER NUMBER (000110.)

DIRECT INQUIRIES TO: Telephone: (215) 563-4100
Facsimile: (215) 563-4044

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

SOLE OR FIRST JOINT INVENTOR	SECOND JOINT INVENTOR (IF ANY)
Full Name Thomas C. Hart	Full Name _____
First Middle Last	First Middle Last
Signature _____	Signature _____
Date Jan 25, 2002	Date _____
Residence Wexford Pennsylvania	Residence _____
City State or Country	City State or Country
Citizenship United States of America	Citizenship _____
Post Office Address:	Post Office Address:
2716 Matterhorn Drive	
Wexford Pennsylvania 15090-7612	
City State or Country Zip Code	City State or Country Zip Code

SEQUENCE LISTING

<110> Hart, Thomas C.

<120> Methods and Compositions for Diagnosing
Palmoplantar Keratodermas and Dysplasias and other
Periodontal Diseases

<130> WFU 99-35

<140> 10/031,889
<141> 2002-01-25

<150> PCT/US00/20400
<151> 2000-07-27

<150> 60/144,644
<151> 1999-07-27

<160> 33

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 45186
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (0)...(0)
<223> n = a, c, t, or g

<400> 1

```
agggagatat aagtgaataa tttggacctg ctctctttga atgtttataa tctggtggaa     60
aaaaaatgga catatgaata ttgatttgtg accagtgcaa aggggggcaaa aattcatatc    120
ccaaagaaaa cggggacaca tcaggtctgt cttgttcatc actgtgtcca cagggcctga    180
cacctagtag gctcagtggg agaaaggagc cccaattacc aacaaaagcc aggaaagaac    240
gggaggctct tacggaaaag ggtgatactt aaactgagca aggaggcacc tggaaatagt    300
gccacctaat aattttttggc gatcagactg acacactaga acggttcata agaccagcct    360
tctcccattg gctagcttcc ttcctcaccc ttctcaccct gggcaagccg cttcctctct    420
ctgggcctct tgcttttcct ctgtaacata aaaggggttg agcaatatca tctctgagag    480
cgccatgtgt gtgcgtgcca gagggaaaac ccccacaacg ctaatacatc aaaactgcag    540
gtttgcacaa aaactgaatt ctgctgaatg caaacaggca aacagcattt accaggaaac    600
aaaacaaaat caagcacata aaaaagtagg aagagttgga aaacggaagg aagataagtt    660
ctcaaacagc tggaatagtt gatgttagct agcgaagttt ttcagaggaa aaaacaagaa    720
gttggttatg aggcaagtgg acctgagaaa aaagactaaa ggggaagaat agcaagtaaa    780
acagaactcc acttgctaga tctctccctc tgtcgcgctc tttcacctga cccactccct    840
tattccccc acacccttc cttctctccc tacgttaccg cacaggaacg aagtctgggt    900
catgtgcgga ccgcttgtgg ctcttaaatc ctcttttgt caccctggcc gtgcaaaatt    960
ttgaaacgtc cctcggcaaa aaaaataaaa ataaaaaaaa aaaatctgtc cctggcctct   1020
tccctagttc tgggtccagt tgcagccaag tgaggggcag cgcgcgctcc caagtccccg   1080
tttcagagac gcgcacgcgc ctggcgccca accccccaatc ccctgctgct cagtgacccc   1140
gcccacgggt ttccgggccg gcgtagctat ttcaaggcgc gcgcctcgtg gtggactcac   1200
cgctagcccg cagcgctcgg cttcctggta attcttcacc tcttttctca gctccctgca   1260
gcatgggtgc tgggccctcc ttgctgctcg ccgccctcct gctgcttctc tccggcgacg   1320
gcgccgtgcg ctgcgacaca cctgccaact gcacctatct tgacctgctg ggcacctggg   1380
tcttccaggt gggctccagc ggttccagc gcgatgtcaa ctgctcggtt atgggtaagc   1440
cgccggctcg gcagtcctcc gggtcgtcct ttctgccctt gagcccctaa cgcagcgcca   1500
cgccaactac cgcttccccc caggcagacg cttgtgggtg gccagagcat cttgactgga   1560
ttcggggacc cttggggacc ttcttccccg ccaggctcgc gaagttaaag ttcatctgct   1620
gagaacttct aactccacac tttcttggtt atcttgggga ctcaacactt tgatcaagaa   1680
ctttttttatt cctcccgctt aattttgttt gctttgagag agacttggga actgcaatcg   1740
```

```
tttggttctc cagtccgatc tggtagcgtt attttaaaa tttatttta ttttttatta    1800
ctatttact agtgaagata gatgagctca gagactctcg aggatatagc atgaagtttt    1860
ctctttttgt tagatggtgg gaaagggact ttctgcccag cgattttggt ttgagcgggt    1920
gttgatgagt actagaaaac ggctagtacc actctgcatt gtttcatgca ttgcaaggag    1980
gtaaaaattt tttaaaaaat taataacaaa gaaaacttaa ctctgaacct agtaattaga    2040
aatgcccaga gtctgcacaa tgtttggctc atggaaggct ctcaataaat acctagtgtt    2100
tgaacatact ggagatattc catatgcctt cagataacat ggttacccct agaacaaaga    2160
acctagtgaa gggggtgggg tggaaaaaag acaatcacta gttagaagag tcacactgtg    2220
gtctacccaa accctcttac agcctgtgac ttctgcagag tcagtaaaaa atcagctata    2280
atttttcttgt caacagaaca aagattaagt catttgttac taaagaaata ccttttaac    2340
tgactgtata gcatttcata atcctgaaca ttgtaacttt ttttttttt tttttgagaa    2400
tggagtcttg ctctgtcacc caggctgtag ttcagtggcg ggatctcggc tcactgcaac    2460
ctccgcctcc cgggtttaag cgattctcct gcctcagtct cccgagtagc taggattacg    2520
ggcacatgcc accacgcccg gctaatattt tgtattttta gtagagacgg gtttcactgt    2580
gttagccagg atggtctcca tctcctgacc tcgtgatccg cccgcctcgg cctcccaaag    2640
tactgggatt acaggcggga gtcaccacgc ctggcctttt tttttttt tttttaaat    2700
ttgagtttga aggttggggc agaaagagat caggatttgc actgccctgt cacatgcaat    2760
ctcccatgtc agagcattac tctcaaacat ggaaaaactt taaaatacac aactctccag    2820
atggacacag ctgaatcatt ttcaggaagg ctctgcctat aaaattacct ttggccttaa    2880
ttcagtaatt aaaggcaccc acaggtctga gccctcatct atgaaggtta aacgtatcat    2940
tcctatcagc actaattggt tttataggat acagacctc agtttgctga ataaactttg    3000
gaaggattca gattatgagg ttctaaactg tagagccttt gaggagaaag gtaccccatt    3060
tcttctccga atagatcatt ttgtgtcttc tcctgggctt agcacattgt cttccttagg    3120
atctgatagt ctgtttattt atcttttgt tgtcatacct tttattcttg catttcccac    3180
ttttgactag cattttgcct tttctccgtt tctggaagcc tgtaattttc aacattcccc    3240
atttctcctt ttgttcagtg gaaaaatttc ttcagtgttg ggatgctctg gaaaggtact    3300
gtagttttgg gggtctccct gcctgctggg ctgcatagca tatctctctt ttttgagaca    3360
aggtctcact gtgttgccca ggttggaatg cagttgtgca atcatagctc actgcagcct    3420
ccaactcctg ggctcaagca atccttccac ctcagcttca gccgctcaaa tagctgggac    3480
tacaggcaat gcaccaccat aaccagcgaa ttttaaaaa ttatttttgt agagacaggg    3540
tctcactatg ttgcccaaac tagtcttgaa gttctgacct cagccccgca aagtgctagg    3600
attacaggcg tgagccactg cttctgcttg cattatctca aatttttaga gcctaacttc    3660
acagtcttgc ctcggagcag gaactgtttg aggcaacaag atggagcctc tggtttcttc    3720
actaggcaga ctgtgctcaa actgggtagc atgaaaggaa agcagcacaa ttaaaattga    3780
aattgggggga ttctttcccc tattgaaatc caaataattt tcctttattg tgcttttttt    3840
ttttttcttt aggaccacaa gaaaaaaaag tagtggtgta ccttcagaag ctggatacag    3900
catatgatga ccttggcaat tctggccatt tcaccatcat ttacaaccaa ggctttgaga    3960
ttgtgttgaa tgactacaag tggtttgcct tttttaaggt tagttttgtt ggaagttgga    4020
tttacatttt caatgatttg atatctgaaa cctcttctga ttagtagacc ctcagaattt    4080
taattttaga ttaagaagat gaccggaatt gacaccactc ttcccaagag tagtagtagt    4140
tggtataatt ttgccacttt atttaaaatt atgatttttag taggtttaca aaataccagt    4200
gctacatttg aatatgtata aattatgttt aaaatttaca tttttggtaag tatgactaaa    4260
ttcttaattt attttcctta ttactaccac tttatttcta aatgttgcca tagtcatttg    4320
gctttgttct aaatctgtag gaaagataga gagattacac attttgtttt cttgcagtta    4380
ctatgctgtc cttcctatca ctacctgttg gctgaggtag tgataggcct aaatgattca    4440
ttatcttaaa tgtactaaat atgttgagta attttttctt ctaaactaac agaaagagag    4500
aacctaggag ttactccgtt aggctggtta aagtgaaagg tagccaagtc aacccagctt    4560
gtttccttct ctcattagga aagaactatt gttcattctc ataacacact ttttccaatt    4620
gcaaacatac tcagggttaa aatagtttag cacaaattgc agcccatttc atttgttctt    4680
cacaagctgg aacttttctt gtaagctaaa tattaaatgg ttcaagtaaa ttggatacat    4740
aagcctgaaa ctaggcgttt ctcattatac atagagtata aattaagaca gacttttttca    4800
tggtgaaagg tttacagcct ttaaaacatc tgggaagaag tgggaaagta gggaataact    4860
ctgttaaata tgataaaaga caaagcacca acaaaggcct agttctaaac ttgttataat    4920
ttctcatggg agttgtggtt tgtcacaagg ttatggcggt ccaagcaagt ttacaatatt    4980
ttttagaata ataactccca gaaatatttt taaaataagg acctttcttt aatatggaaa    5040
aaaaaaagat gaaatagaga ggaagaggtt gccttcctc aactggattt gctggtgtga    5100
tatgcagctg gtggaaacca gtttgtccat gctagctgtt gatggccacc agtccactga    5160
attagtggaa acccatgctt aaccagttgt taactatttt tagtatcatt ccagatgct    5220
acgcattttt tttgtaaaaa acatacttac aatagtatct atataaattt ttaaaattgt    5280
aatgtatatg tggtatattt aacctgaaaa taatctttgg tgtatacgtt gatgagccta    5340
ggcctttgca gacctcttac agttactctg tcacagccct tcaaaggttc tttgacttca    5400
gattaagaat caattgcatg tggaatgcat gtgcaaagga agagatattc aaggcaattg    5460
ttaccatgta ccataaacct tgtacataat ttttgtcat tttctttcc tctgtctcta    5520
tcccttcttc tttgacacac aatagacacc tagtcaattt attacaaaaa aaatgaatga    5580
atgaagtgga attcagttgg gaaataggtt aaataaatta tttaggagat gaggaatagg    5640
```

```
taaaaagaga caagtacata gtttattctt ttgacttaga aaacttttga ttcttaaatt    5700
ctgcagaatt ggagaaactg gtggggaaac ttctaaaatc attatttaat taccagagat    5760
gtaatagata tagacaaaag cagttttctt cttttattat tttttcatca gttagttctt    5820
agcttaaata gtagtccaaa gctggtaggg acagagggaa ttagctggtg gctgaatgag    5880
gaattgtatc actttttgtg aatcacggtg taagcacatt tggtgttttg ccattgccta    5940
agaacattag tcacattagg tcaatagaaa atcacttttt aaagccaaat aaagttatat    6000
gtgttcccaa ccatgagttg gaaagaatta atatatatgc tgttggaggg tagaaccctg    6060
cctaatcata tggttctgga tggcattgat cgaatcctta ttctttcatt aggaataaca    6120
atagaaaaaa tactcctgcc ctactgattt caggatatgt ctattttaaa gtgcccattt    6180
gacaaaacca ttatcagggc catgttttct ttttctgcag aaaaatcaac cactctggtc    6240
agtagttagg tcttatgaca agcaccataa tttccttagg cagagtagaa tataataggа    6300
tacttctttt tgaaacttaa tataatcagg tagttccaga taaacatagc ttgcaaagtg    6360
ataaaatacc atgttatttt agtaaatcca attgcaagag tgatgggaaa cagagtttaa    6420
aaacttaaga aagatattaa gatggagttg actttgaata ataaagtcat ccactgttga    6480
tgggtgacat ttattaatac agaaagttta cagatttтac cataagcatc agggtatttc    6540
ctgcagctgg ggaaaccgtg cttgaaagga tgcgtaactc aaggaaaaca caagcccatg    6600
taataagtat tgcatgtgag aattgttcct atagaattag aaagcatctt ttacattaaa    6660
atttatttt gtaaaaagag aaaacatcaa aacttgagta gtatttgcta ttcaaagagt    6720
cttacataaa cgaaaacata ctcaacctac tgccattaca gaaatatttg acaaattctt    6780
gccatgctta cctgccatcg ttgttgttat cctacaaatc aattggattt tcacgcctct    6840
ccactgactg gaaccctaca acttgcttcc ttttatcctc tttatatatg cttcagatat    6900
gcttgaagta gattgtttct tattgttctt gctgctcgag tttgttgagt agttggtaca    6960
tgcagagtat ttgtatgtta tgacatatat aggttttgtc catggttcct ggctcataac    7020
tcactcccaa gacccttgtt acagaaacca gaatctctct ctctgatctt ctcctaccct    7080
cctttcatct gcccactgca gaactctaat ctgattatgg tttctaagac cctcatacca    7140
gagagtattc tgccccatac catagcagaa ggaacactgc acagagacac caagaagaat    7200
ctgaacagac aggccttgtt aggtttagat catgtcctta taacctaatt atatttaac    7260
atggttatcc atgctttaat catgtgtatt caatgaacgc tccataaaag cccaagaaga    7320
acagatttga gggaattctg aagcgctaaa catgtagggt ctgacaggaa ggtgaagaag    7380
aactcatcat gctggaaggg tggtccaccc taactccgca gggacagaag ctcctgtgtc    7440
tgggaccctt ccagaacttg ctctatgtat ctcttcaagt gcctgtgtat ttatatcctt    7500
tcaaatatcc tttgtaataa atcataaaca tgtttccctg agtttttatg ccactctgtc    7560
aaattaattg aacccaaaga gggggttatg ggaaccccaa tttgaagcca gtccatcaga    7620
agttctggag gcctggactt gagactggtg tctaaagtgg gcaggcagtc ttggggacta    7680
agccctcaac ctacgggatc tgaaactgtc tcctggtagg tagcattgga gttgtactgg    7740
agggcactga gctggtgtct gctgcagaat tgattgcttg cttgctggtg gggagaactt    7800
cctacatatt ttggggtcac cgaagtcttt tgtattgatt gttgttgctg agctcctatt    7860
gctagaaact tgctttatta agttctgttt attcagggct tatctgaaag agaaacattt    7920
tttatgattt gagatttcta agccatttta aaacttggtt ttaacagctt gagaaattgg    7980
gggtatagta gggatggaga tacatatatt tggatgtgac ttcaagctag ataaaagttt    8040
ggaaggaata aaagtttgac atctaaagtt tttgcatagt ttgagttgag caggaggtac    8100
taggtatgtt tttaaaatat tttttcagc caggcacggt gactcatacc tgtaatccca    8160
gcactttggg aggccgagac aggcgatcac ctgaggtcag gcgttcgaga ccagtctggc    8220
caacatgaag aaacccgtc tctactaaaa atacaaaaat tagctgggtg cggtgacaca    8280
tgcctgtaat cccagctatt caggaggctg aggcaggaga attgcttgaa cccaggaggc    8340
agaggttgca gtgagccgag atcacaccat tgcactccag cctgggtgtc tcaaaataat    8400
aataataata aaaaaatgaa agattttttc ttactcagca tcctccaggc attttattat    8460
ctgagcactt tatgggagtt gcatattaca tttagggccc actcaggtgg gtgggtatct    8520
aagcatttga aataacctta tgtaaactaa taaggagtaa tcaggcctgt ggcaagatgg    8580
aaacagtctt agaggcattc aaattcaaat ttcctttaaa acactgggct ggccaaaaca    8640
aaagacaata ctatctacag gccagtttct aagattatca gattttagta gcatttacca    8700
tttcattgta cttggcacac tttagcaaat ttgcacttct taaaagtacc tgcaggcaat    8760
ctcctatata aaaacacaat gcaggctagc ttggctcctg cctttaattc cagcactttg    8820
agaagttgga gactagcgtg gccaacgtgg tgaaacctca tctccactaa aaatacaaaa    8880
attagccagg catagcagcg cacgcctgta gtcccagtta cttgggaggc cgaggcagta    8940
gaatcacttg aacgctggag acagagcatg gagtgagctg agattgcacc actgcactcc    9000
atcctgggtg acagcgtgag actctgtctc caaacaaaac aaaacacaca cacacacaat    9060
gtaacaacac gaaacagaat actgtgaaaa tgcttaatta tgtctgactt tacatgatgg    9120
ctggatatgt gattattttt tcttctttat gctcttatgt actttgttca tttttaatga    9180
tgatcatgta taaagctcct ctgtgtagca ttctctccac caaattgccc agagacagga    9240
agtcctgtaa aacaaactaa gctccaaaaa atgacctcct gttgaatagg cttttttttt    9300
ttttttttt ttttttttg aaatggagtc tagctctgtc tcccaggccc tcgctccttc    9360
cacctcctgg gtttaagaga ttgtccttcc tcagcctcca gagtagattg gattacaggt    9420
gcccgccatc acgcccagct aatttttgta ttttagtag agatgggggtt tcaccatgtt    9480
gaccagactg gtcttgaact cctgaccca agtgatccgc ccgcctggac ctccccaagt    9540
```

```
gctgggatta caggtgtgag ccaccacgcc cagcctgaat aggctttcta acctactgtt   9600
tctcatttta ctttctctga ggcagtaaaa gaaactgact ctaaaaggga gcagtagaga   9660
aggaactcag attttatttt gaagattaag ctactcaagg gctgaggcaa tatgtagagg   9720
ggaagtagac tcattatggc tggaaatttt atttggtgat agtgaagcaa atcttagggc   9780
tttctaattg agctcttgtt tgaaaggctc caatcttaat agaactataa gctaaaaaaa   9840
atgaccatca gtgtttctaa agcaagttgc tactcaaaac aagaacactt tgggaggctg   9900
gggctggtgg atcacctgag gtcagaagat ggagtccagc ctcaacatgg tgaaacccct   9960
ctatactaaa aatacaaaaa gtagccgggt gtggtggtgc acgcctatag tctcagccac  10020
ctgggaggct ggggcagaag aatcgcttga acccgggagg tggaggttgc agtgagctga  10080
cattgtgcca ctgcatttca gcctgggtga cagagtgaga ctctgtctca aaaacaaaac  10140
aaacagaaac aagaactcta tgttgagaaa tccatactag agggtttaat ttctcacttt  10200
gtgtgctagt gattaataaa tttcaagctt atcacacagc cagaaatgtc gcctgtgttt  10260
ctacaataaa agatttggga atattgtgac attttttcac tgagccttct gggttcatta  10320
ttaagatata acaattttag aagacttatt gagataggta tactttttaa aatttggatt  10380
caatatatca agctcagcat tttgtctttt ttttgttttg ttttaaacaa ctttgggttt  10440
acttgtaaga gtatttcagt ggaggaaaag tatcagaagt ttagcactga gtgcctgcct  10500
aactatttta tgtcttcttt catgatgatg cttaatctca acataatgac tcagtttacc  10560
attgtaaact ctttgtagga tgtcactgat tttatcagtc atttgttcat gcagctggga  10620
actgtgggga tatatgattt gcccacatctg aggaacaaac tgggtgagct ccattgaaaa  10680
atttctaact ttcttattaa aagcaagctt cctcacttcc tgatggctta tgcaaattaa  10740
tgcctattta ttcatcagag ggtgactgga gagattatgg ttttcttatg aatttcttgc  10800
ttgagcctgc ttgtctgctg tttctaagtt gtctgcacta ttttatgtga gtaatttttct  10860
actttattat gttctgtttt cacatgtcaa atcagctctc ccaagaatgc tacttgtaac  10920
ctaagtagac gtgaaccgaa aagggtaaag acccagctaa aaaaaaggtt gactcaagtt  10980
cagttacat tcgtaaagta gttaggctgc cttaggttgc ttattctttt cttgaaaaga  11040
atccctcaag agagaaacat gtgaggccac agcagcttag atctgtcttc cacagagaag  11100
gtggctttta cagagaaatt gacatactca tcacttatct gacatgaccc agctttgtaa  11160
aactggcttc tattaaaata gtcttaatag actattcact gaggaggaga gaatcttatt  11220
cactctttac attctcttca cattttcaga atagatgttt aaatcattgc tcacactgga  11280
tccataaatg tctaaaatgt tgatgaagaa ataggtgatt agagagtaaa attaatagag  11340
acttaccctt ccttgcattt taacataata ttctttcccc ttttccttcc tctgttactt  11400
ggctcttaaa taccagaagt gagatatgaa aaagggaact gggaacaagt attgaaagca  11460
ccataggttt atcttatatt agcattttcc aaactttata atgaaccagc agtgacccac  11520
acatctgcca ggtaggaatt gctcatcagt tgtgtcgcgt tatcttgtta agcttcagca  11580
atgttccata cagccactaa taacagatca aagtgagcat tagggttgaa attagtaagc  11640
tgtttcttct cagttctttc tggtagttga ataaatatag aatgtattaa atagttttct  11700
ttatttcaga acttctcaga gtctgtaata tattgtatgg tggtagctta ggaggaagat  11760
gcaataggaa acttttccca gataggttca ctatttttt tttccacgaa aaataagctg  11820
ttctcaaaat acagtttaca aaattttatc cttaactctt cactctttct cctagttagg  11880
gagaccgctc caccagtaga aaagataaac cctggtaatt tgttgtgtaa atgggataaa  11940
tagcctagta cctagtcatg tggattcagg cagcactgag cctaaattaa agtttgcaag  12000
gtatacatgt taatgtatct aagttactat atttagcctg tttcttaagt atgtttcaga  12060
aacatattcg ttttttttcag tggcagttac cttcagatgc atgtgcttct aaagcatgtt  12120
ggttgcatgt gcagtacatg ttgcttaagc atttagcttc agaatggcat ctttcctgtg  12180
aatgtcttaa catttacaaa aatataccag gatctcaaat atcagtgctg gtattttttt  12240
ttttttttac ttaaagaaac tgatatgatt aaatattaag agacaatatg atccttgttg  12300
gcttgtaacc ctagttttta ttgtcttgta gttattaaat agagcatctg ttgagggact  12360
cttttaaaac cacagccatg aacagacgtt ggggctaaga gacagagcag cctgcgacag  12420
tgtggaccta cctgtagcag ctagcaaagg cctctagcag ctacagtccc ttctggagtc  12480
tttatttgca tgcaaaatgc aaaggagtcc tggtgaccta cctccaaggc agctgccctc  12540
ctgaacactc ccttggaaaa cagtaaacat cattttggaa tgtgaacaac cagagactac  12600
acaggagaaa ggaaaaaaaa attctgaaga tgcaaaatct tgggtggctt caccgttcag  12660
ttttttaata aaaggaacaa tatacaacac gttgttcttt ttctcttttg aaatcccttc  12720
tattacagtg attttttttct aagattgtca ggatttgaag tgtatgtttt gttttattca  12780
cagcgtaaat tttattcaca gtttaactgt ctgcctgagt gtctttcctt tctctaatta  12840
ccttgaggaa cccaagagcc tgttgtaagg agaaaataag gcccttggat ctcttgagat  12900
tcacagatat aagttattga agggaagatg gtcctatgga ggacatattt aaagaaggga  12960
aaaaagaggc tttctcagat atgtcagact gctatagtat acttctacag attatagacc  13020
tccagtacct ctggccagaa agatggtatc gtaaacaccc tattttttt cttttctttt  13080
ttcattaggt acaagctttg tgctaagaag ttgacatact ataagctaca aaagtctgt  13140
aaagtagata taactagttt cattttatag atagagaaaa ttaatctctt acagtgctaa  13200
gctcacagag tttctaactg taaaatgcta gaacttgtct ttcaagccta aagacttcct  13260
tggggctaaa tagtgaaaaa agccatttca caaataagta aatggtattt agaggcatat  13320
ttggatttcc tggtaaattc cagtctgtga gcatcatgaa tattagttta atgttgcatg  13380
ggctcatgtt gaagtttttaa gagaagaact gccttgaagc ttaggttttcc ttagctatta  13440
```

```
ggctactgac tttcttgcct aaaccagggt tttttcattg aagaccaaaa cttaccttct    13500
ccttcagttt gtagtttgga aattggtaga agagctttgt aaacttcaaa ttaagtacaa    13560
actaagtgtc atagtcaaat ttactaatct taattacagt attgttcaac tgattgctat    13620
cttctagctc tttcctgccg aataatggtc ttgtttcctg ctctgttggt ttagagctga    13680
cttctttcag ctttggtaag cctgaaatta tggggttatg tttaattcat attgtctggg    13740
tggactttcc tctcttgcat ttctgcttga atagaagaat ttttctctag agagtagttt    13800
gtcatcctta ctctgttgat tcagatgact ctttgtatga tctgagaggt atactgttct    13860
gctattctga gaagaagtat ttcagaaaga tgaattaaga gtacagtgga ctgctcccac    13920
ctggaaactt ttatctatct cacctctgga cctgataaat tctttatcac tcaggacctt    13980
gatgacgctg ctctctgaaa ccctccccag ctctctctat taccgtgaga aacatcagaa    14040
cttтggttcc cattgcatat cgcaggtacc tctgctttca tgccatgctg taatggagtg    14100
attgggtagc atgttttcat ctctttccag attgaaaatc tgtatttctc cctgtatatc    14160
ttcaacacct aatgcacata gaactttgta ggtacctgga aaatgcacca cagttttctt    14220
ttcttttttgc agacttttca caagtattac caacttacaa agaattaatt ttgtaggatt    14280
ctagaaagac aaatcaggaa tggtgccata tacatctttt ttgattccct gctctaaaga    14340
atattatcag gttaccttcc tgcagagttt taaaagaatt gcatatttca agctgacttt    14400
caggatgtaa atataaccaa agcaactgat atgtaaaaaa tatattcaat ggcattccta    14460
gattttcttc tagggtgttt tattgttttg ggttttacat ttaagtcttt aatccatctt    14520
gagttaattt ttgtataggt ataagaaagg ggtccagttt taatttctg cgtatggcta    14580
gccagttctc ccagcaccat ttattaaata gggaatcctt tccctattgt ttgttttтgt    14640
acggtttgtc aaagattaga tggttgtaga tgtgtggtct tatttctgag atcttcattc    14700
tcttccactg gtctatgtgt ctgttttтgt accatgcttt tttggttact gtagccttgt    14760
agtatagtat gaaagatagc atgatgcctc caggtttgtt cttttтgctt aggattgtct    14820
tggctatacg agctttttтt tggttctata tgaattttaa aatagtttct tctaattgtg    14880
tgaagaatgt taatggtagt ttaatgggaa tagcattgaa tctgtgaatt gctttgggca    14940
gtatggccat tttcatgata ttgattcttc ctatccatga gcatgtaacg ttтtтccctt    15000
cgtttgtgtc ctctctcatt tccttgagta gtggtttgta gttctccttg aagagatcct    15060
tcacttcttc tgtattccta gatattttat tctctctgta gctattggga atgggagttc    15120
attcatgatt ttgctctctg cttgcctttt gttggtgtat agggatcctg gtgacttctg    15180
cacattgatt ttgtatcctg agactttacc gaagttgctt atcagcttaa gaagcttttg    15240
ggctgagatg atggggtttt ctagatatag gatcatgtta tcttcaaaca aagacaattt    15300
gacttcctct cttcctattt gagtacgctt tatttctttc tcttgcctga ttgccctggc    15360
cagaactccc aatactatat tgaataagaa tggtgagaga gggcatcctt gtcttgtgcc    15420
agttttcacg gggaatgctt ccagcttttg cccattcagt atgatattat ctgtgggttt    15480
ctcataaaaa gctcttatta tttgagatac gttccttcaa tacctagttt attgagagtt    15540
tttaacatga agcgatgttg aattgtatcg aaggcctttt ctgtgtctat tgagataatc    15600
atgtggtttt tgtctttagt tctgtttatg tgatgaatga cgtttattga tttgcatatg    15660
ttgaaccggc cttgcatcct ggggatgaag ccaacttgac tgtggtagat aagcttttgg    15720
atgtgctgct ggatttggtt tatcagtatt tcattgagat ttтttgcgtc gaagttcatc    15780
agggatattg gactgaagtt ttcttтtтgt tgtcgtatct ctgccaggtt ttggtatcag    15840
gatgatgctg gcctcataaa atgagttagg gaggagtccc tcctttтcaa ttgtttggaa    15900
tagtttcaga agaaagggta tcagctcctc tttgtacctc tggtagaatt caactgtaaa    15960
tccatctggt cctggacttt ttttcattag taggctattt attactgcct cactttcata    16020
acttgttatt gatctattca gggatccaac ttcttcctga ttcagtcttg ggagtgtgta    16080
tgcatccagg aatttatcca tttcttctag attttctagt ttctttgcat agaggtgttt    16140
gtagtatttg ctgttggttg tttgtacttc tgtgagatca gtggtggtat cctgtttatc    16200
attttttatt gtgtctgttt gattcttctc ttatttttga caagctgac aaaagaagc    16260
aataggggaa ggactctcta ttcaattaat cctactgtat atctggctag ccatatgcag    16320
aaaattgaaa ctgttcctgt ttcttaatct atatacgaaa atcaacttac gatggattaa    16380
agacttaaat gtaaaaccca aaattataaa accctggaat agaatatagg caatatcatt    16440
ctggacatag gaatgggcaa agatttтatg agaaagacac caaaagcaat tacaacaaaa    16500
gcaaaaattg gcaaatgaga tctaattaaa ctaaagagct ctgcacagca aaagaaacta    16560
ctgtcagagt gaacaggcaa ccaacagaat gggagaaaat tttttcaatc tatccatatg    16620
acaaaggtct aacatccaga atctacaagg aacttaacaa atttacaaga aaaaaggagc    16680
cccattaaaa agttggcaaa gaacatgaac agacacttcc cagaagatat tcatgtggcc    16740
aataaacatg aagaaaagct caacatcact gaccattaga gacgtgcata tcaaaatcac    16800
aatgagatac catctcatgt cacaatggtg attattaaaa agtcaaacaa catgctagtg    16860
aggttgtaga gaaataagaa cgctttтaca ctgttggtgg gaatgtcaac taattcaacc    16920
actgtggaag acagtgtggt gattcctcaa ggatttagaa ccagaaatat cattactgca    16980
tatagaccca aaggaataga aatcattcta ttacaaagat acatgcacat gtatgtttat    17040
tacagcacta ttcacaatag caaagacatg gaatcaaccc aaatgctcat cagtgataga    17100
ctggaaaaag agaatgtgga acataaacac catggaatac tatgcagcaa taaaaaggaa    17160
tgagatcctg tcctttтcag ggacatggat ggagttggaa gctgttatcc tcagcaaact    17220
aatgcaggaa cagaaaacca accaccacat gttctcactt ataagtggga gctgaacaat    17280
agaacacatg ggcacaggga ggggaataac acacactggg gccagtcagg gggtgggggg    17340
```

```
tcaagctgag ggagagcatt agaaaaaata gctaatgcat tctgggctta acccatttat   17400
gcctagtgtt ccatttctgg aatgctaagc atgtggaagt tctttatatc ctgctcaagg   17460
tcattgccaa ggtctgattt ttcacattca acaaattgca acctctggca taaatgggtt   17520
aatacctagg tgatgagttg ataggtgcag gaaccacca tggcacatgt ttatctatgt   17580
aagaaacctg cacatcctac acatgtaccc tggaacttaa aaaatttaaa atatatatgt   17640
atatatattt aatatggaat tttaaaaatt actaatgagt tcttttatct gagtaatttt   17700
gcatcaacat gcttttatta tggaagagaa gattcagtga gtacaaaatt gcagatacat   17760
gtgtcagaag atccctgaat ataataaggc ttagtattct gtgtcataat tgcctgtttg   17820
tattcctctc tggtctttaa acttcattag ggcaaggatc aactccatct tactaaccat   17880
ttgattccct atgtattaca cgatatatga ccaataataa gccttcaata aatacttgta   17940
aaataaagaa tgttatgtaa tatatcatgt ggtattgttt tattgatgtg ttctttgaag   18000
tgttaccttt gtgtccttag aggctatggt gggccctcag gccatatatg ttggatgatt   18060
tagtggaaaa gagattagac ttttatattt tattctattt atttattttt gttttaatct   18120
tgtgtaagcc tatagaggcc aagatagggc tcagagttaa aaacagtgat gttctaactt   18180
tcaaaactac taaaaacaaa attagctatc aagggaaatc ctgagtgttc tgactgcagg   18240
ggttccccaa agcctggagg accgaagctt tgaaatgctg catggtgtgc ctcacctctg   18300
ttgagtatag tgaggaacat gagctgagga gtggattaga tgaactctta agattttttt   18360
ccaactttac ggtttgagtt tatgagtaag taagtgagga atgagcccag ggactttgac   18420
tactttgctg tgcagtaaat tgctgttttc ttagctaatt ctggagtgtg tattttggtt   18480
ttaatgtaat gtcctttcag cactggaaat aaacagattt atgaccctcc attttcacat   18540
gtaacttcca gttccttcag taaaatgctg gaagttgctt ggctgtaacc actgtaacta   18600
tttcctggca gtttttgcatt ttattacaat tcgacttttt aaatgcatgg ttcagtccag   18660
ttaataggca gaaaactata gctattcctt ggaactgaca tagggaagtt ggtttctttc   18720
tgaggaaaga tgtgacactc tgatggattt ggcaaaatac caattcagta aaatttgcct   18780
ggttcctttc tgacagcaaa tggaggtaat tgtacgtata tagatttcta cagtaaacaa   18840
agggcaaaaa tggaaacatt ttttgttttt taatgttgct ggaatataat atttccatat   18900
ttggttaatc atcaaatcat tcaactccat gcaaagtact ctttttcaaaa ttacaaatca   18960
tcttgatgga ttagacttgg atgtctaaat tttttattat gtattctaaa ttatctatag   19020
gtacattttg aaaattatat atacatgtat cactgtacta catttgttat aaaatgtgaa   19080
atattaagcg attaaatact ttattaatgt cataaaattg tgccctaccc aaaatatatt   19140
ttaagaagat atcatttgaa aataatgtaa aatattcttc attattatga tttttgttgcc   19200
aaaagtaatg ttagagtccc tgttactgat aattttcatg actgtaaatt ttcaacattt   19260
ggaaacaaaa ttttctgatg tgttttaaaa tcttaaagaa tttcagtaga ttttgacaca   19320
tttttgttaa aagtatgaaa gtgtcacagc tttagaagga tttagaataa ttatcaattg   19380
gtattaggga tgacaatttt aatagacaaa ttttaactga aacatttgct ttattgatgt   19440
ctaggactaa aaaaaatatt atagtgatgg aaccatttca aaacatctgt ttttaaaaac   19500
ttctgtcaac atgtttgaca gaagtcagaa tagggtaatt ataaaaaaac tgccattttc   19560
ttatgtttgc atcattagta tttgcaaatc cttgtttctt tgcaagtata tttaaaacta   19620
cttgttcatt ttatgagtgt tagcctgatt aattcagata agatgctcaa gtcagtttaa   19680
gcaggcagat gtgaaaaagg gaatatgctg tgccttaact atcttatagg gacttactag   19740
cacactgtac ataacagatg aacatctgat attagatctg atgggcacac caatgttgtt   19800
tcatatatta aataatagtg gaagttttca atttaattag tttatttatt tatttatttta   19860
tttttgagcc ggtgtttccc tcttttttgcc aaggctgggg tgcaatggca catggttgca   19920
gctcactgca atctccacct cccaggttca aaggattctt ctgcctcatc ctcccaagta   19980
gctgggatta caggcgtgtg ccactatgcc aggctaattt ttgtattatt agtagagaca   20040
aggtttcacc atgttggtga ggctgctctc aaaatcccga cttcaggtga tccacccgcc   20100
tcagcctcct aaagtgttgg gattacaggc atgagccact gcgcctggcc aattgtattt   20160
ttatgttaga aatatatatg ggagttacaa tactttgttt ccaatgttcc gctggaaacc   20220
actggttgaa aggataggtc ccagtgtcag ctgcccagtt ggcagttttg gatagaatat   20280
aagttctatg aaggcatgga tgaagtatcc ttcatggtaa tgatggtaat gttagtatag   20340
tcacttgttt aatgttgtat aactttttaa ctcttccaac ttcaatttca tctatatctt   20400
tcatagcagt ccagtaaagt agtatcaatt atagatatca ttcctacttt cagataaagt   20460
aactgaggct cactagtttt gacttgctca aggttacgta aatgataagt gacagactgg   20520
atttaaaggg cggagtccac tttactaaac tgcctctgtc ttctagccaa gtttgttatg   20580
taaagtaagt attcagacat tggctaagta tattattaat caagaaaaca gtagaacagg   20640
tttaaggaat tttttttttcc ctctcctcag tgtcaaaaag cagtcaagta aaagatccaa   20700
gataaataac ttcagttccc aaaaataact tcaattccaa aatttcccta tgactaggta   20760
tatctaattg gatagctgat aaaatgagtg ggaaattgaa gaagggatat tttgagaatg   20820
tgtggcaact taatggtaat attgaagaag caaatggata gttattgaat cataatagtt   20880
ttttaaactt aaatactagg gcttgatctt catattttag gaaattatct atttgttgca   20940
gtagaaatta gcataaatga atggaaactt aaattatctc atctgtgcaa ggcagtcttc   21000
agagactgat gtttatataa gtgaaaactc agaaaaattt tttggctgca tttcaagctg   21060
gatagcgggg tcaaaacttc aacttcttaa ttggtgatac aacgtcagaa atcacaaaag   21120
aagtatttca caggaaaaaa tgacataccc cgaaaacatt caaatcaaga aaatcacagg   21180
ggtacagtgg atggtccccc ttaaggtact gcgtacttgg cttccccgct ggtaaatgcc   21240
```

```
tctcttttac taagttgttg tcgtgcagag tggcattgca gctctgtact tttacagaat    21300
aatttcaggt tctaaaattg atacattatt tcagattgta agaatagtat aaaataaaat    21360
tttaaaaggg aaaataattc catcaattta atggaatgtg tagggtttaa gttataacaa    21420
caaaaacaaa gttgtagttt tttaggatta ctcatataaa tagtgtgtct attaagaatt    21480
actggctttta tgaatattaa ataagaaggc tggccgtggt ggatcatgcc tgtaatccca    21540
gcactttggg aggccgaggc agacggatca cctgaggtca ggagttcaag accagcctgg    21600
ccaacatggc gaaaccctgt ctctactaaa aatacaaaaa ttagccgggc gtggtgacgc    21660
atgcctgtaa tcccagctac tcaggcggct gaggcaggag aattgcttga acccggaaga    21720
cggaggttgc agtgagctga gatcgcgcca ctgcactcca gcctgggtga cagaacaaga    21780
ctccatcttg aaaaaaaaaa aaaattaaga caaccacaga tataaggatt tgcaattaga    21840
aagcttgaag ggaaattttt ggaacttcta ggtcaaccta aaattacaaa tagggaagca    21900
gctttggaga ggaacatttg cacttctgaa gtaatactgt gatttggact gtcagacatt    21960
gtgatatatt ggcatcacca cacctctgtt aattataaac ttacctattc tcagtgcctt    22020
gttaactgtt gaaataggac ttgacatttg ttgttcttcc tcttctgtta ttcatttatt    22080
cactcattta gtagttaatt acttcctagt gccagatagt tttctaggta ctgaataaac    22140
aaaaatccta ttgtcatagt tatgtttttg tggaaagaac aacctgatag actatcgtga    22200
aaaataaatg agttagtata cataaaacac agcccatgag atatagtctg taattatcat    22260
ccctgcttca tttatttatt tatttatttta tttatttaga gaatggatct gattctgtca    22320
tccaggccgg agttcagtgg ctggatcatt gctcactgta acctcaaaac acctggcctc    22380
aagtgattcc cccaacctcg gcgttccaaa gtgctgggat taaaggcatg agccactgta    22440
cctggccatt gctcctgttt taaagatgag gaaactgaaa gtacagagag gaacgtgact    22500
tgctcaggat cacacagcca atcagtggca gagcagtcta ggcagtctag gcctgaaggc    22560
attattcttt ctttctgttc tgcgtcaaaa accctaggca ggcatggtaa aaagactgaa    22620
agagccaagg caacataggt acctgaagaa tgggagaaca cctatgagga ctgctagaaa    22680
ttttaggcag cccctttaaag gccttgggca acaatttgaa cgtgactcat gagtcagtta    22740
attctctgca ccttttttttt tttagagaat acaatggata aagcaatgac ctttgtaaca    22800
tgtaaacctg tatttgaatc ctttcctctc catttaattt tgcagagtca ccttttcttc    22860
attagaaaga aattttatcc caaagagatt atcgtgaggc tcaaataata taggtggaag    22920
cccttttgggt ttaaaaaaaa gccttttctg tcatttctgg gtatctcccg tcagtgattt    22980
tagtaattgt ttacctctct ggctcttcgt gagaagacaa attgtaggaa gaagtggagg    23040
atggcattgg gtaggatgct tagttctagt tccaccactt ccttaggtct tttttttgtc    23100
ccccaagact atatcatttc tgtgttctct tgaacctgta agatttatga gtaatctgta    23160
caatatttat aaagtatgaa atctctgcca ttgtggcaca aaagtagcca tagacaatgt    23220
ataaatgaat gggcatgctg ttctccagta aaactatatt tatagaaata ggcagtcttc    23280
tccaatttta gtttgctcac cccttttccta tagaactcag gtttatcata tacaatgtta    23340
cacattattc ctttgaataa attgtttacc taatactttt taaatttggg cattttctag    23400
gacattctat tgtttttatt tttaacttgt gttatactgc agtgaatgtc ttcctgtgtg    23460
tagcattttg tttgtgtatg tgtgtcaata tctataaagt agctcttaag aactctggaa    23520
acattgtcag actgcttcct caaacagcag aggtgggagg gagcatcggc ctaaccctac    23580
ctggcctgcc ttaccagcat gcaaccaagg agaagagaat aggatcaaga gccagtgcca    23640
gagagaagta cataatacaa ctttatggag tttaggttgc atcctataaa gttaccatgt    23700
tacactgttt cgatctacca aagtagaagt tttacatagt ttagtatgaa gttaatttta    23760
ttaacctttt cacacaacac accctctcat attgctgccc aacaatgtcc tgggtctttg    23820
tttccatggt catcatattg agtagggttt tagcagggac ttgagtagtt gggaactgaa    23880
ctcatgctag tgaagccatt cttcactaga agcagacata tgggatccag aatatcatga    23940
acaaccagct gggtacagtg gctcacacct gtaatcccag cactttggga ggctgaagga    24000
agcagatcac ctcaggtcag gagtttgaga ccagcctggc caacatggtg aaaccctatc    24060
tctactaaaa atacaaaaat tagccaggcc tggtagcaca tgcctgtagt cgcagctact    24120
tgggaggctg aggcaggaga atcgcttgaa cctgggaggc ggaggttgca gtgagccgac    24180
atctcgccag tgtactccag cctgggcaac agagtgagac tctgtctgaa aaaaagagaa    24240
tatcatgaat atcaccaaag ataaaagcaa gatgcggata tgagtagccc agagtaactt    24300
ttcctgtgct atataacaca gtgagtggtg attcccacaa ttcaagtcca gtttttctgg    24360
attctaaagc catacatttt taactctccc cctaaatatg atttgaattc agagagtcag    24420
atgcaataag gtaattttaa ctttaagtgt agtttagccc acacatttct ctaaactggg    24480
ggttggcaaa ctacaaacct cttttttgtaa gtaaagtttt attggaacac agcctagctc    24540
attcatttat gtattccctt tgcacctttt aggttgagta gttgtgactg aggtcttcca    24600
gccttcaagc ctaaaatatt tactatctag tccttcaaaa agaagtttgc caacctctgc    24660
aaacaagtga aataatgtgt attagagtag agcaggagtc cacaaactac aacctctgga    24720
gccaaaacca ccctactgtc tgttttttgta aacagaaatt ttatagaaac atagttgacc    24780
cctgtgccaa tttgtttgag tcttctttt gatggctttt gtgctacaag ggcagagttg    24840
aacagttgca agagaactaa gtggcctgca aagcctagtg gtctttgctg tctggccatt    24900
ataggaagtg tttaccaaac cctgcatttg atcatggaat cccagaatgt ttataccttt    24960
gaggattttt ttttctactt tccatatttt acaggcaaca gaactaaagc ctcgtaaaat    25020
cttgcccaag gttctacaaa tagtctttttc tattaaatta agtcagagcc tgaaattgtc    25080
tgtgcccaag tctctcctca gaaattagc acagcatttt cccttttctgt ctcccacatg    25140
```

```
ctacacagtc atcctggatc atgtgccccg cctccatgag gagatgttct aacatgaggg   25200
tgactaggct ctggagccag agcatcttag ttcaaatcct ggatctgtta aaaataatag   25260
ttattttttca gttctaacac tgaaataaga caaaaggcca tattactgaa ataaagtgta   25320
aagaatatat ttaatactca ggttacttta gcccctcat gatcagaata tgctaataaa   25380
ttgagagatt gacagaaaaa aaggattctg tttcactgtg gctcacacct gtaatcccat   25440
cactttggga ggccgaggtg ggtggatcac ttgtctggag ttcaagacca gcctggccaa   25500
catggcaaaa ccccgtctct actaaacaat acaaaaatta gccaggcgtg gtggtaagca   25560
cctttaatcc cagctactcg ggaggttgag gcaagagaat cgcttgaacc tgggaggtgc   25620
aagttgcagt gagccaagat cacaccactg cactccagcc tgggcaacac agcaagactc   25680
catccctgga aaaaaaaagt tcatcatttg ttcttgtagt ttcctagtgt ggcccttgtc   25740
tcagtggaaa attgtagtgt gatagactga ggagaattct gtttctgcca cttcgcctgt   25800
tgccttggta ggagctcctt atctctgagc tgagctcctt ttgtccttgg taaaatgtgg   25860
ttattatcat ctgcatcctc atagtcttct tgtaaggact gtgaagtcac ttttgaaaaa   25920
tgccttgaat atgccagatg gtctgttttc tttctttttt gttttgtttt gttttgtttt   25980
cccttggcct tcattcctgg aagtgttttg ttatatacat tttgtttggt ggccaacagt   26040
gctaaccaca gaaaagtttg tatgtcttct ttttctagaa taattcttac agtgtaacct   26100
ctcagggtga agttttttgtt ttgaaagggg cacatttact gtgaatgaga gccatggaaa   26160
tggacctgaa atttgaagac agattttact tggttttctt ttctcttatt tgtttgcagt   26220
ataaagaaga gggcagcaag gtgaccactt actgcaacga gacaatgact gggtgggtgc   26280
atgatgtgtt gggccggaac tgggctgtt tcaccggaaa gaaggtggga actgcctctg   26340
agaatgtgta tgtcaacaca gcacacctta agaattctca ggaaaagtga gttgctacaa   26400
atgagacata cgtttagttt tgttttatag tatatatgaa tatgtgtgta cattttttgga   26460
atttttagttt gattatacaa aatatctttg gcttagaaat attaggcatg ctatgtaaaa   26520
ccttactgga aaaataaatt gaccaacatt attgagagta ttttttcaaa gtgttccaaa   26580
agtaatggac caatgattac tttaaatgaa atcatgtaat ggaccacaga attgcaaatt   26640
actaataaag aaaagccatt ttgcttattg ccatgtaata acatgttgca tgatacaagt   26700
agatacgtat gtttatgctg caacaagtat aggtgatact aattgggcaa cttttaaaca   26760
agaccataaa taactgaaat caaagttctt agtatttatg cagcctgttg gtttgcgagg   26820
gctgccataa caaagtacca cagactgggt ggcttaagga aggttgtttt ttcacagttt   26880
tggaggctag aagtccaaga ttgaggtgtt gtcaggtttg gtttcttcca aggccacact   26940
ccttggcttg cagatgtctg cttgcttact gggctcctaa acagtctttt ctgattgtcc   27000
taatgtcctc ttcttataag ggcatcagtc atattggatt ggggcccacc catatgacct   27060
aattttacct taatgacctc tttaaagctc tgtctccaaa tacagtcata tggctgtgat   27120
acaggggggtt agaatttcaa cgtatgaatt ttgagggggga cacaattcag actataaact   27180
gcagttaatg tttactgtta aattaaatct ataactgact aaaccacaca aacaggagat   27240
ttcaaattag actttattag ttttggaagg aagagatgaa atgttgttac ttttctgtt   27300
ttatgtagtt gtaaaagcca ctaaaaatgt acatgtacaa atcatcccaa gccaggcaac   27360
attagtaatg aatgactgca gcacagaggt aagagagatg atttaaagga ggatgaattg   27420
tctgcaaagg tgctgggcaa gatctagact aattcatccc cttcatttta agtttaggtt   27480
ttaaatagct ttgtttggct tgatcctaga gctacacatt tactttaac ttgtttactt   27540
tgtaccctat catttaggat atgctatgta ctattgtact ctatttgata tttcaaacat   27600
tctctcattt agtgaagaag ctcccaaccg gagtgctcag aacctacatt gcctcacctg   27660
tgaagtgagc atgtgggact gaattgcctt tgaggtctct ttcagctctt aatgatctgt   27720
tatcccatag taagatacat tatttttaat ctcgttggat cttaaacacc aaaaataata   27780
gtatttaaga cgtaggatgc tatcttgtca ttatttaat gcacatgtca acacacgagg   27840
ttttgctaga tgttatgata aagcaagcga aacagcaggc tactgctcct caaatatcta   27900
cagtcaatga aatattgact ggcgttgtgg gaaaatatct caaaagtatt ttatttacaa   27960
tttaaaatat attctttttgc ctaaagaaac aatgaaacaa ctataacttg ttttttgttat   28020
tgtgttagtg tagtcataca tatgatacag ttttataatc tgcttgataa gggtacgtct   28080
ctgggtttttt cacactttgc aacaagtggg aaagaacttc gacaaaaaca ttaaaacatt   28140
atcaacctgg ttgcacgttt ttctgattat tttcagtgct cttaggtttt tagaccatat   28200
caaaattctc tctctacaca catgatttat aagggaaaga aaagtttgca gtatgcctaa   28260
aatgttcctc aggttaatta tatcttgttt tgcatagttg aatgtttatg gaagtcttag   28320
atagtgagtc actctgaaac cacagactct ttttttaatag taataatatg gtgagttcta   28380
taattcctag catctagcac acaggtatga taaatgtttt aattaagtga ataatcagct   28440
tcctgaattt ttctactttg taaatatgct ctttgtgtgt tgaatctatt tacatatata   28500
tccaagcatt tggcacaagg taaaccaatg tggctcagat tttaatgtta ttgtgaatca   28560
cttggggatc ttgtttaaaa aaaaaaaaaa aaaggagatt ctgattcagg tgatctgggc   28620
tgagtactta cagtatgcat ttctaataaa cttctgggtg ttgctactga tattggttca   28680
agaactaaat tctgagtaac aagtataaac cgaatgtcca taaagtataa tttctttttag   28740
tatcatatgg ttataatgct tgcattttgt atatatgttg catacaggtt ttgctttct   28800
cttttttctt attaaaaagg ttataaaggt acactgtaga aactggtaggg agtaggggaa   28860
agcttggcat ctcatgccac cattcagagg taaccatgtt tggcattttta aaaatgcatt   28920
tatagtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgc gcgcatacat atatatgtat   28980
gtattataca tatacgttaa tatacatata atatacatat attactgtgc aaaagttgat   29040
```

```
aaattgatta tgattatgta taaactttga tatcttctca tatttaatat aatatagttg   29100
tgagcactct tccatatcat taaataactt tctagacatt ttaaatgaat atatagaaaa   29160
tgaaacatca ttaatcagtc ccttatctta gaatatttat ttcggttatt tacaatctat   29220
tagtagttta aaagtacttc atggaaaatc cttatacatg catcttagtt aacttgaact   29280
tttatttta gagaagacat gccaggctgt tctaggctat tgatacatat tgccaattac   29340
cttcctgaaa gattgtacca ctttccactt aggcacagag tgtgaatgcc tgttttaatt   29400
actctattta catcagtaaa gactatatct gaaattatct catttttggg gacaggtatt   29460
ctaataggct ctacaagtat gatcacaact ttgtgaaagc tatcaatgcc attcagaagt   29520
cttggactgc aactacatac atggaatatg agactcttac cctgggagat atgattagga   29580
gaagtggtgg ccacagtcga aaaatcccaa ggtaatcaag cacacatttt atcattaata   29640
aaaatatgaa tgctgaatac catcctccct cctaagcagt cctaccagtg ttgctcgcca   29700
ttttattggt catcccagtt gtaacttttt atgtgatctg ttaacgatct ttatctccta   29760
ttgatacata ttctgtcact ctccaagtcc tatttatgag ttttttctttc acatgtctgt   29820
tgtttccatt gcccccttttc cattttctac actctggcat gcagctttgt cctgtttccc   29880
ccatgctatg gattagcctc ctgccggtaa ccacctgagg ggttcttcct gcctgctgca   29940
taaagaaaaa ccatggcaat atagtagaga aagagcttaa gagacactag gctggccatg   30000
ccacgtgggg gatggagttc atattcaaat tatcttgtcc aaagctcata ggtaggggtt   30060
tttcaaaggc agcttggggg aaggggtggg ggtggccagg taacagatgc ttgctgctga   30120
ttggttgggg tgggtgaaat cacagggagt tgaagctgtc ctcctgtggg ctgaattgct   30180
tctaggtggg gccataggag tggggttgct gggtccaggt agaaccacgg gtgtcagaca   30240
tgcaaaaata aaataaaata aaataagata anggtaagat aaaataaaat aaaataaaat   30300
aataaataaa ataaatataa aatttccctg aaaagatatt tcaaaaagcc agtcttagat   30360
tctacaataa tgatgttatt tgctggagta attgatggag ttgcatgtct tataacctct   30420
ggaataacgg ctgacaatct ctcaagtctg cgccttagct ggactcaggt tcctcttctc   30480
cccacagcct gactgcctcc attagcttca caaaagtggt tgggtttcag ggcaaggccc   30540
attgtcattt aaactgtagc cgaaatgact tccaaagtta gcttggccca atagcccagg   30600
aatatttaag tggaaggcaa gatggggggat gggttagctt agctctcttt cactctcata   30660
gttttctcac tggtataatt tttgcaaagg cggtttcatg cctgccatct ctttcgtcgc   30720
tacctctccc agttcccatt cttagctgtt ttatgaaatg cttctagttt catcctctta   30780
taccaagttc tgggagactg atttgagtaa taataaaact ccagtttccc atacagccgg   30840
ctctgcgtga attaaactca ttttctattg caatttccct gtcttgataa tcagttctgt   30900
gtaggccgtg aggaaggaga acccgttggg tgattacgaa actgtgttac tgcccactac   30960
ctaatggata catttagcct ggtatccaaa cccatctaat tatgaccata actatattta   31020
tcaccttgct ctgctttcaa acgatatgac acacaatgaa tgaaaacttt catttttcat   31080
cttcatttgt gctgttccct ttgcctcaaa tagccctcaa cttgcctaca gtaactgtaa   31140
aatttgccac ctaaaaaaaa atctcaaaat cctctctatg ctttgatgtc cagcaaaaaa   31200
aaaaaaaaaa aaaaaaaagt tatcagtata gctctattac tctcaaccag aggtggcatt   31260
tttcatcttt acccataagc cccatggtat atttcttagt ttgggaaatt atatcataat   31320
actttgaatc tgtctagtca gaaatttaga taatttaaaa aaaaattatt tttctgaaag   31380
tatagctgga tctagcttag gatctacatc tatctaatat agttccttaa catnttgtta   31440
aatggccacn ggnataagtc ctgtaatgcc atactttgct ttaggatcat gtgactaagg   31500
ggtaaggaat tggaaagcaa tggggagcta gcagaatttg aaaaattata taggtaggtt   31560
attttttcctt aatacattga aatagcctca aattctcaga gaatacaatg tttaatcctc   31620
taatatctgt agagttcatg gctcaaattt gtattatttg aataaactac taatagatta   31680
ataattactc aactaaaaca ctttgaaatg tgagagttcc cttatctccc tcgcaggcat   31740
atgacagagg tgtggcttct cacacctttg gttgcccttg ccctaccacc caaacccta   31800
gggggagcat gcagaggggc aggtgcagag gccatgggag gtgctttttgg gctctggccc   31860
cacagcagtg tctaggagtg gatgttggag actcctgaag cccaagtggg catgtgttac   31920
agtgtgctct ttcagcttag ccgtctgcag atggcttgtg ttaatcaggt cattagaccc   31980
catgccttat tgcaagggca ggggtccaat gtgacagcct aagttcttgc tcagtgtacc   32040
agaagaattg gatcacacgt gggctggaag gatgagcaca aggtttttatt gagtggtgga   32100
ggtggctctc cgcgagacga ctctcagcca gagaaggagg atggagggga aaggtgttct   32160
tccctggag tccacttgtt cctgttctcc tctgggttca aatacctctt ccctcctttt   32220
ctgctgtgct gtcccaccac tctccacagc tctgtgccgc tctgttcctc tgctcctctg   32280
gatgttcagc tccttgtatc tgtgcctgct aagattttgg gtttatatgg gggaaggatg   32340
gggggcatgg cgggccaaaa ggcacctttt ttggtgtgaa aacagaaatg cctgtcttct   32400
cttagggcct taggtcttca ggcttgaggg tggggccttt gctgagggac caccctcttc   32460
tacccagtat ttccctgtct cctgtccata tcaacattac taactttttc atctgcagac   32520
taataatgct aaggtgtggc atttttcaac tgtgagacta tgtgaaggtt tttcctgtcc   32580
aactgatggc atcctcccat aattctaccc ctttcttaaa agaatcttct gcagtatttc   32640
tccaagtttta ttctagagaa tttctttgtct gtgaaatgct ctagttaatc aaaattggaa   32700
aacggagcat atcatatccc cttctcaaat tcaccaaagt gaagtcctaa tgtgtcttaa   32760
tgtatctgca tgagacagga agctgagatc tattcaacaa caaaaatcca aacaagcatc   32820
aagaggagga gtgttagcac ttgagcctag ggagactgtg gctcctgcct gaaagatggg   32880
agcctcagtc acagctgctt taccaagtgt catatgctat gtttctgagg actcctgcta   32940
```

```
aagctccctt ctccctccag ccaaccactt ttgttttaga caagggctgg gtttatgaag    33000
gactgttttc atgactaaag ctttatagaa ggtttaagat aaggagatgg aattgagtga    33060
agtaggaaat atgaaagcag atattataat ctggcttcct gatttttcac tagcattttt    33120
gtttataaat tagttctgtt ctaagaatcc aatgacgtaa tagaaactct caaagattct    33180
taacttgaga tatagggagt ctttgaaact gctgaaatta cagacagcat ttattgttta    33240
tgagcatttc tgaatctaga gctttcacta gatttgtaaa gaatgtgggc caaaagatta    33300
agagccaatc ctgtatcttg tactcaaaat gtttgtaatt ctcacctttt tatcccccag    33360
actcctcttg ctctctcttc attttcacag tgttactgga aaggggtcct gatccagacc    33420
ccaagagagg gctcttggat cttgtgcaag aaagaattta gggcgagtcc tcaatgcaaa    33480
gtgaaagcaa gtttattaag aaagtaaagg aataaaagaa tggctactcc atagacagag    33540
cagccctgag ggcagctggt tgtccatttt tatggttatt tcttgattgt atgctaaaca    33600
agggatggat tattcatgcc tccccttttt agaccataga gggtaacttc ctgacattgt    33660
catggcattc ttttttttttt ttgacagagt ctcactgtgt tgtccaggct ggagtgcggt    33720
aaagcaatct tggctcactg caacctctgc ctcctgggtt taaatgattc tcctgcctca    33780
gcctcccaag tagctgggat tacaggcatg acccaccatg cccagccaat ttttgtattt    33840
ttagttgaaa cagggttttg ccatgttggc caggctggtc ttgaactcct gacttcaggt    33900
gatctgccta ccttggcctc ccaaagtgct gcaattacag atgtgagcca ccaaacctgg    33960
cctgtcatgg cattcttaaa ctgtcatggt gctggtggga gtgtagcagt gaggaagacc    34020
agaggtcact ctcatcgcca tcttggtttt ggtgggtttt agctggcttc tttactgcag    34080
cctgtctgtt ctatcagcaa ggtctttatg acctgaatct tgtgctgacc tcctatctca    34140
ttctatgact tagaatgcct taactgtctg ggaatgcagc ccagtaggtc tcagcctcat    34200
cttacccagc tcctattcaa gatggagttg ctctgggtca aacacctttg acaacatcat    34260
taagcctcag ttctcacact gtttttgttt tgtttgtgtg tgcaatgatg ggcaaatctc    34320
tgccttatag gatggtagaa aaaaggaact taatattgta atgactgttt tgtgccagat    34380
aattgcttta aactgtatca gcatcttatt tagtcctgtt aatgatatga atgttatctt    34440
cataacagct gccattttat taaggactta tcagagaaaa acactgttct aagcacttgt    34500
tacccattat tacattgaat tttcataaca acccctttgag gtaagcatga ttatacccac    34560
tttaatagaa gtgaactgta gtttgtaat gttaggttcc ttgccaaatg ttacacagat    34620
agtaagtgat aaaatcatat gccctgaaat tacattatgc tgccaaactt aaatttcttt .  34680
tttatccttt atattagtat attcttaggt ttaaacaaga caactagtta acacatacta    34740
gattttgtcc acagttcctg gctcataact cccatagccc ttgtcactat cttttaancg    34800
ttggggcaca ttaggcctca gaagtaggcc tcangaaaac agagtctctc tctctctctc    34860
tctctgatct tcttccgccc tcctctcacc tgcccaggcc agcactttaa tcttctcctg    34920
cctttctgat cttgggtcat aagaccttca tttccaaaga tgtcctgtgt cataccctaa    34980
aggaaggaac actgaacaga gagaggctca gaagaatctg gacaggcctt gctgtgttta    35040
catcattccc tttatgtcca gtcacatctc tacatggttg tcagttgtgc ctatttgatg    35100
aagtccccat ataaggctca caaggacagg gtgcagagag cttccagata gctgaacaag    35160
tggaagttcc tggagggtgg cgtgttcagg gagggcatgg aagctgtgtg cccctcccc    35220
cataccttgc cctactcatt tcttcatctg tttcatttgt agtatctttt ataataaacc    35280
actaaacatt agttagtatt tctctgagtt ctgtgagtca ctctagcaaa ttaattgaac    35340
ccaaggaggg tgtcatagga tccccnacat tatagctggt tggccagaag cacaggtaaa    35400
caacctaggg ctttcaattg gcatgagaag taggggggcag ttttgtggga cggagccctc    35460
agcctgtgag atctgatgcc atctctaagt acacagtgtc aaaactggat tggaggacac    35520
ccagctagta ttcactgtga aattggttgc ttgcttgatt tgtggggaaa aacccacatg    35580
catttgatca cagaagtctt ttgtgttgac agttgatagt gttcagtgag agaattaaaa    35640
aaaaattgag tttcttcttc aacatactct ctcaatgtga aaccacagaa actatttcca    35700
ttcaaagatg gaaatggttt gtttgtcatct tagttttat ttatacatct tagaagaaat    35760
gtccaagctt tgttttttct ctcaccctat atataaaatt acctatgagg cacagatttt    35820
tatgatcctt gattatatag actttgtcca aattgtgtgt tttatagcat tactgtaact    35880
tgttatagta atctttgtgt atattatgtc tcttaacatt gtcttccata ttgttaatga    35940
ccatctcata tttatctctg tatcatgtat atcttcaacc aatgtgactg gcttaggaga    36000
aaaaattagt gaacaattaa ctagtgtttg tgtaatctat acaattgtca aggttacaat    36060
tgctattttt gaagaaatcg ttgttgtttt tctctttgtt tcatctcagt tccattttgt    36120
caaggattcc ttttttttttt ttttttttttt ttttgaggc ggagtcttgc tctgtcaccc    36180
gggctggagt gcagtggtgc aatctcggct cgctgcaagc tccacctcct gggttcatgc    36240
cattctcctg cctcagcctc ccgagtagct gggactacag gcacccgcca ccacgcccag    36300
ctaattttttt tttgtatttt tagtagggac ggggtttcac tgtgttagcc aggatggtct    36360
caatctcctg acctcgtgat ccgcccgcct cggcctccca atgctgggat tacaggcgtg    36420
agccaccgcg cgtggccct tgtcatgtat tcttaacctg tgttatatct aagagaaggt    36480
gtgagaggcg ggaccatttg tggataggtg cagagggcat gaagtagccc agaaaactaga    36540
tttgccattg actaaatttt agcctagtaa aaaacatgtg gtaaccactg aaaaactaga    36600
acagtgtgat atgacaattgc ccagttgaat aagaaattca aatagtttat agtaacaaaa    36660
aataatttttt accacattgt gactagtgcc ctaaaaacaa atttgaagga gcagagagag    36720
agaggaagtg agttttcgct ggggggctggt gagtggaggc tactttatga gcagtttga    36780
aaattacaag ttagggaaaa actcttaggg gagtttttaag tggtgagcaa caggcacttg    36840
```

```
aggttacagg agcagcagca taaagatata gaacagagag gccatttggc agtcttgggg   36900
aaactccaag aagtcgaatg tgtctaaagc tgtggttctc tctctgtgtg tgtgtgtgtg   36960
tgtgtgtgtg cgcgcgtgtg tttgtgtgtg tgtgcatgtg tatgtatgtg tgaaatcgct   37020
aatacagaga agctgttaga acaattttct cacttcccaa aaatgtccgt tcattcatgg   37080
gcctaggcaa ctctcctttg tgtgtcttac agcccaacac tgtcatataa ggtgtggttc   37140
tataatgaaa cagttttatc tgttgtttac ggagctagtc agcctgtgtt atgcttgtta   37200
ctggttgagg gtgtccaggt tcttggcttc ttgaacaaag aattggacaa aactcacaaa   37260
tgaggcaagg aaagaatgaa gcaacaaaag cggagattta ttgaaaatga aagcacactc   37320
cacaggatgg gagtgggcct aagtaaatga ttcaaggccc tggttacaga attttctggg   37380
gtttcaatac cgtctagagg tttcccattg gttacctggt gtatgcctta tgtaaatgaa   37440
gagggtggag ttaagttaca aagtcattta ctcagtatag gccttgtgtt aatggagagg   37500
gtgttactcc tgggggttgt ggcccatgta aacggagagg atgaagtgaa gtgacaaagc   37560
ccttcgcatt cctgccattg ctgaagtgtt tccactttat ttagttctag gaagtcagtg   37620
tgaattggcc ttatgttccc tgcctccaga acctgttctc ctgcctcaca tagacccttt   37680
tttccctttg ccaacttcac ttcttttaca gccacctcac taccaatgtg tctgtctcct   37740
aagtcaaaca tcagtgattc ctctgttttc tctaaaccct tcttatgtct tctacttctc   37800
atctctttct tggtcaaaaa tctttcaaaa atgagtaaga atgcagctat tcaggcaaac   37860
taaaaataac atcacagtga tatacaaaac cagtgtcatt tcacaaagga aaattatcaa   37920
tactagatcc tgaaaaagaa acagcgaatg aaagccattt acacaactcc attgtgtaat   37980
tgacacattg aatcactcat aaaacaggtg ctctgggtct gaatctagat cctagctagt   38040
ctggtagctg aaatcataga attatagtag agtttaggaa atcatcctca aaggaaagat   38100
tatatgttga tatcaaatgt atatttcctt tctaggccca aacctgcacc actgactgct   38160
gaaatacagc aaaagatttt gcatttgcca acatcttggg actggagaaa tgttcatggt   38220
atcaattttg tcagtcctgt tcgaaaccaa ggtaaaaaaa taagcctaag ttttttgtta   38280
atttgtttgg aactatttat tgaacagttg ctctgtgtga tggatttcgg ggatacctag   38340
atggaatggg catgatcccc tttttacaga aatagaaaat aggtggccta tgaattattc   38400
tttcctttta tatccatgac aaactttagt aaaaaatttt cttttctact gagcttagca   38460
tttattcagt actcttctca atatattttc caggtagcta gtgacaatca gagtgatatg   38520
taagacaaac tcatttgtcc tcctagtagg aaagattcta gtagaagcaa agaattgtgt   38580
accattctgc aagtggtttg ttggaatctt tctttgatac ctgttcctgt attcccccac   38640
ccccattaat ttagtcatta attactacat gaacccgtaa aataaatcct taaattattt   38700
tcctggtaga ttttttgagc tgtgtaagga cctttcaatt cactttacat tagaatacga   38760
tatgtgatgg taagtattaa cccagcttcc tgagtgatgg cgtgagggcg gaacccaggt   38820
tcatgaaaag tattcctatt attgaatttt accagttatt ttcagaagtt tgatacaatg   38880
tgagttgatt tcataacatg tcctctaact gcacttgata gcaaattatc ttgttatcct   38940
gagttgtagc caacaatgac ttggagaatc tatgcaatac tcagttttat tacttctaag   39000
ctcattttga agataatact acccatggct gatttgttac tataaaatag gtttagtatt   39060
tgctgtctgg aaacattctg attagtgtct cctgggagga ttataaattt atagtaccca   39120
aagantaaac ntgttgtttc cctttcctaa acttttagtg nataatncag tctctgccgt   39180
gtctcatttt catcacttgc cctccanagc tcccatctca ctgaattctt gcagtgttct   39240
gaatgttgag agccccaang tgggtcttat aacagccagt cagcaacatt tctgtttttc   39300
atctgacacc aagggtctcg tctctttgct tttctaccag ttattctggg ctcttcagct   39360
ctaaagaaag tataggtcct gaaatctttc cctaccttct caatttcctg gggagggctt   39420
ctttggaaag tgggattgga aataagataa atttgaagat aattgagaaa tgaatggaaa   39480
gtgaaattga agggtccatg ttaagagatt gcaagttatg ctatcaccaa atagattttt   39540
tgtgcctgag aggattaatt catagtgcat attatgtgtt gactttatca ttgaggtcct   39600
ggcacatgat agcattggca tgatataatt tgagctactg atactatagt gttgcttctg   39660
gtgttgttac cagaacatct aaaatatatt aggatttttt taatggcaga ggaaatgaag   39720
actaatatga catagtcctt gtcctagtga tttacagttt agcaagacat acaagcaaaa   39780
cattaaagta aagcatgata actactttaa taaagcattt tttaaattca ttggtaacat   39840
aagagaaggt agaagagttt agcaaaccct tcccaaaaga aatgattgac aagttatatg   39900
agataataat tcagggaaag gaaattcgac tttctaaagc caaattattt gacattggtt   39960
ttcatatagc ttggtaaaag ctgttatttt ctcccatgtt ctttattctt tgactgttat   40020
aagtatgatt tgtacagaga aatggcaatt tcaaaacaga gggctttgat ggattaattg   40080
ctttgaattg atccctcatc tacagtatct tgtcaggtac ttgagaaaat aatgtactta   40140
aagtttcctc ttttgacttt cttttggtat tctatactgt aagttgggga aaaaagtatt   40200
ttctcttcct gctaattggg ctacttgaaa attcccacca actttgccaa taccagtgtt   40260
ctgtataacc cagaattcag aattagctgc aattaaggga attcacagct tttctgtagt   40320
cagagagcaa attgaaatta aagaaaaaag aaatagtggg aggacaaatg aggttttacc   40380
tttacacttg aaaacagatt taagaacaag ccttatacct agatttatta atactttgag   40440
gtatgagagg gaagagaagc ttagaaatac ggcagaatgg gctttctttg ttcttctccc   40500
agctatgctg tttttatttta ttattgtatt tttaagagag aagggaagtg tctctcctgg   40560
gtcacattaa ttaggaaata cagagtgttt tcataatgcg taaagtctag tccatttaag   40620
tcttgtttca aaatgctatt tatattattt gagcaggaag gcagagacct taaactgctc   40680
cacccaattc attttacaca aaagattaaa aagaaaaaaa cagtgtcaaa aggtcaagtg   40740
```

```
cccagtgttg cacaactaat gataggcaaa acaagagagg aggggaaaaa aaagaatcct   40800
gactcccgac tcctagttca gtgtgttctt cactgttaga aggtgctgct gaacatagtt   40860
ataccatatg aagatcacac tatctatttg agattgtaga aatttgatta ctgcagagct   40920
ctggctgggc tagattcact tcttattctt tattgcattg cagtattctt aagagataaa   40980
tggctctttt agaatcagat tggccttggc tgttgaaatg aggcattaat taccttggta   41040
gctgacacat tcttacaggt caggggcttg atgaagttta tcttcttccc ttgttcctgt   41100
gattgctctg taaatagaac acattcagag cccttgaatg cacttgcagc ctgtgcctcc   41160
cacagtgatc gatggtcaga taatgggagt ttaatgacca gtactgagag agattatttc   41220
catggctgct atgggccaag aaagctgggt ggtcagaaag ggacctttc cagactctcc   41280
tgggttgtgt tatttctttg acattggttt cctttcatga ggcgccaagg tgtatttgta   41340
agttgtccag tgttgcacag ctaatgaggg gcaaaacaag aaaggaggaa aaaaaaagta   41400
tcctcactcc tagttcagtg tgttcttcac tgttggaagg tactgctgaa cacagttata   41460
tgttcagtgt acaatgtaca atgttctttg acattccgtc atttgaagga tgggtcctat   41520
ggatagtatc atctgcatgg ttttgaaaac aaaagatata agctaatttt gccctgtcta   41580
gtgactacga gacagggaga gaaaatctga atatttgtta aagtagacac agacccataa   41640
attgaaaagg acactaatcc tgccttagga gacagtaagg cacttgtccc tgctatctat   41700
tagctgtgtg tccttggaca gctcattgct tctttctgag ccaccgttac acgttatcta   41760
tcaagtgagc aggttgtaca ctagatgaat tcacaggtcc ttcttccaag tgcttttcta   41820
atcttcataa tttagataat ctctcagtag caaaacagtg tacaatatgg tcaatctgag   41880
atttttaggg ggagaatttt agggaacatc agaaatggca gtagttaaaa ggaaatagga   41940
ctcacaggct gacttctctc tataacttca catggtagaa gggatgagag ttctctctgg   42000
agccttttgt attaggtcac taatcccaaa ggccccacct cctaagacca tcacttttgg   42060
gattgggggtt ttaacatgat tttagaggga cataaacctt catctcattg cattagagat   42120
atttgaaagc tcagctcacg tgtattcctc cacagctcac atgtattcct ccatcaccca   42180
acctgatggc tttgaacgtt gtagacataa atcctttcat cattatcaag aatattgcca   42240
aaagcttctc agaaattatg aggggttttt ttagtttcta aaatattctc aaagaaagtc   42300
ccatgtacta atgtttgcct tttgatgaaa aaggatgaaa tcttaatgat tgccttaata   42360
agctcaacaa tgcttgttag ttgagtcttc ttattgtgct gattcttata aacaacaaca   42420
ttcagtataa acattaatgt atgtgattca ctaaggtttt tgcatgattc tctgtgaggc   42480
ttcagatgtc tcttggatta tgtgtcttttt tttcatcgcc agcatcctgt ggcagctgct   42540
actcatttgc ttctatgggt atgctagaag cgagaatccg tatactaacc aacaattctc   42600
agaccccaat cctaagccct caggaggttg tgtcttgtag ccagtatgct caaggtaagt   42660
gttgcatttc agacaccatt tatgagctat ttacctgtgt gcagctggct gttgttggca   42720
aaggcaaaag gatgatgcag tagagagagc gcagtgtcta tagtcagaaa atctgagtgc   42780
aagtctggcc ctatcactta ttaatggatg attgctcatg gaatttactg taccatccag   42840
caaaatgtca atagttacta tatattgagt gagctctgct tgttacatat attgcctaac   42900
aatgctcaaa actctgagaa gtagtaagta taatccctat ttatgggcgg ggaacaggaa   42960
ctaagaaatt tttctaataa tttgaaggtc tcacagcttt tagcattgga gtttcacttc   43020
taatcatcgt ctccaaaacc caactttat taaaactata ctaacactgg tttctctctg   43080
ggagaatttt aaaattctgt acttagggct gggcacagtg gcttatgcct ataaccctat   43140
cactttggga ggctgagatg ggtgaatctc ttgagtccta gagtttgaga ccagcctggg   43200
caacacggcg aaacccottc tctattaaaa atacaaaaaa ttagctgggc gtggtggtgt   43260
gtgcttgtag tcccagctat tcaggaggcg gaggtgtaag aatcacctga gcccaggagg   43320
tcaaggctgc agtgagccga tatcatgcca ctgcactcca gcctgggcaa acggagtgag   43380
gccctgttat gaaaaaaaaa aaatctgtac ttaggctttc agatcaggct gtatgtgatg   43440
tatgtcgaaa acacagctat aattgattga gggagaaacg ttaccatttt aaagtttatg   43500
ctttcaagcc cagatttggc cactaggaat ttcccagctc actagtgaaa ctgctgatga   43560
gtgattattt gccagtgagc ctttcattct ttctaaaata tgtactacta gttgtgactt   43620
gtaggctata ggggctataa tatatcaaga caatctttat cctcatgaag cttacagtta   43680
agtaagagat agagattaaa taattataac aacagagtga agaacagtga agaaaaagta   43740
cagagttata atatatataa tagggccagg actgcatgag gaaggtagga aagacatttc   43800
ggcaagaggt tgtcagggaa aagacttgct tgagaaagag ccaagttgtg gggtctggct   43860
gcttagcaat gaccataata cctaactttt gctattttta catgaagtaa ctaatttaac   43920
cctatgagga aagtactact accatctaga ttttacaggt aagtaagcag agatacagag   43980
aagttaaact cttcacacgg ctttggcttt aaacctatat aggcttcaga gcctccccac   44040
ttaaccactt tgccatagct acatccatat taggtgctaa gtagatatct gttaagtaga   44100
aggaggatga aaggatagtt agctagttgg aaaatggatg gatgaatgaa gtgatgctta   44160
agctaagaac aactttcagg ggtaacatgc aaagaataat ggagcaaaga agaaaaaata   44220
gaaaatggga taatcctttt tctactaagg ggtaaccatg tgtgttattc atcttcaggc   44280
tgtgaaggcg gcttcccata ccttattgca ggaaagtacg cccaagattt tgggctggtg   44340
gaagaagctt gcttcccta cacaggcact gattctccat gcaaaatgaa ggaagactgc   44400
tttcgttatt actcctctga gtaccactat gtaggaggtt tctatggagg ctgcaatgaa   44460
gccctgatga agcttgagtt ggtccatcat gggcccatgg cagttgcttt tgaagtatat   44520
gatgacttcc tccactacaa aaagggggatc taccaccaca ctggtctaag agaccctttc   44580
aaccccttttg agctgactaa tcatgctgtt ctgcttgtgg gctatggcac tgactcagcc   44640
```

```
tctgggatgg attactggat tgttaaaaac agctggggca ccggctgggg tgagaatggc    44700
tacttccgga tccgcagagg aactgatgag tgtgcaattg agagcatagc agtggcagcc    44760
acaccaattc ctaaattgta gggtatgcct tccagtattt cataatgatc tgcatcagtt    44820
gtaaagggga attggtatat tcacagactg tagactttca gcagcaatct cagaagctta    44880
caaatagatt tccatgaaga tatttgtctt cagaattaaa actgccctta attttaatat    44940
acctttcaat cggccactgg ccattttttt ctaagtattc aattaagtgg gaattttctg    45000
gaagatggtc agctatgaag taatagagtt tgcttaatca tttgtaattc aaacatgcta    45060
tattttttaa aatcaatgtg aaaacataga cttattttta aattgtacca atcacaagaa    45120
aataatggca ataattatca aaactttaa aatagatgct catatttta aaataaagtt    45180
ttaaaa                                                               45186
```

```
<210> 2
<211> 1838
<212> DNA
<213> Homo sapiens

<400> 2
aattcttcac ctctttctc agctccctgc agcatgggtg ctgggccctc cttgctgctc     60
gccgccctcc tgctgcttct ctccggcgac ggcgccgtgc gctgcgacac acctgccaac    120
tgcacctatc ttgacctgct gggcacctgg gtcttccagg tgggctccag cggttcccag    180
cgcgatgtca actgctcggt tatgggacca caagaaaaaa aagtagtggt gtaccttcag    240
aagctggata cagcatatga tgaccttggc aattctggcc atttcaccat catttacaac    300
caaggctttg agattgtgtt gaatgactac aagtggtttg ccttttttaa gtataaagaa    360
gagggcagca aggtgaccac ttactgcaac gagacaatga ctgggtgggt gcatgatgtg    420
ttgggccgga ctgggcttg tttcaccgga aagaaggtgg gaactgcctc tgagaatgtg    480
tatgtcaaca cagcacacct taagaattct caggaaaagt attctaatag gctctacaag    540
tatgatcaca ctttgtgaa agctatcaat gccattcaga agtcttggac tgcaactaca    600
tacatggaat atgagactct taccctggga gatatgatta ggagaagtgg tggccacagt    660
cgaaaaatcc caaggcccaa acctgcacca ctgactgctg aaatacagca aaagattttg    720
catttgccaa catcttggga ctggagaaat gttcatggta tcaatttgt cagtcctgtt    780
cgaaaccaag catcctgtgg cagctgctac tcatttgctt ctatgggtat gctagaagcg    840
agaatccgta tactaaccaa caattctcag accccaatcc taagccctca ggaggttgtg    900
tcttgtagcc agtatgctca aggctgtgaa ggcggcttcc cataccttat tgcaggaaag    960
tacgcccaag attttgggct ggtggaagaa gcttgcttcc cctacacagg cactgattct   1020
ccatgcaaaa tgaaggaaga ctgctttcgt tattactcct ctgagtacca ctatgtagga   1080
ggtttctatg gaggctgcaa tgaagccctg atgaagcttg agttggtcca tcatgggccc   1140
atggcagttg ctttgaagt atatgatgac ttcctccact acaaaaaggg gatctaccac   1200
cacactggtc taagagaccc tttcaacccc tttgagctga ctaatcatgc tgttctgctt   1260
gtgggctatg cactgactc agcctctggg atggattact ggattgttaa aaacagctgg   1320
ggcaccggct ggggtgagaa tggctacttc cggatccgca gaggaactga tgagtgtgca   1380
attgagagca tagcagtggc agccacacca attcctaaat tgtagggtat gccttccagt   1440
atttcataat gatctgcatc agttgtaaag gggaattggt atattcacag actgtagact   1500
ttcagcagca atctcagaag cttacaaata gatttccatg aagatatttg tcttcagaat   1560
taaaactgcc cttaatttta atatacccttt caatcggcca ctggccattt ttttctaagt   1620
attcaattaa gtgggaattt tctggaagat ggtcagctat gaagtaatag agtttgctta   1680
atcatttgta attcaaacat gctatatttt ttaaaatcaa tgtgaaaaca tagacttatt   1740
tttaaattgt accaatcaca agaaaataat ggcaataatt atcaaaactt ttaaaataga   1800
tgctcatatt tttaaaataa agttttaaaa ataactgc                          1838
```

```
<210> 3
<211> 463
<212> PRT
<213> Homo sapiens

<400> 3
```

```
Met Gly Ala Gly Pro Ser Leu Leu Leu Ala Ala Leu Leu Leu Leu Leu
1               5                   10                  15
Ser Gly Asp Gly Ala Val Arg Cys Asp Thr Pro Ala Asn Cys Thr Tyr
            20                  25                  30
Leu Asp Leu Leu Gly Thr Trp Val Phe Gln Val Gly Ser Ser Gly Ser
        35                  40                  45
Gln Arg Asp Val Asn Cys Ser Val Met Gly Pro Gln Glu Lys Lys Val
    50                  55                  60
```

```
Val Val Tyr Leu Gln Lys Leu Asp Thr Ala Tyr Asp Asp Leu Gly Asn
65               70              75                          80
Ser Gly His Phe Thr Ile Ile Tyr Asn Gln Gly Phe Glu Ile Val Leu
                85              90              95
Asn Asp Tyr Lys Trp Phe Ala Phe Phe Lys Tyr Lys Glu Glu Gly Ser
            100             105             110
Lys Val Thr Thr Tyr Cys Asn Glu Thr Met Thr Gly Trp Val His Asp
        115             120             125
Val Leu Gly Arg Asn Trp Ala Cys Phe Thr Gly Lys Lys Val Gly Thr
    130             135             140
Ala Ser Glu Asn Val Tyr Val Asn Thr Ala His Leu Lys Asn Ser Gln
145             150             155                         160
Glu Lys Tyr Ser Asn Arg Leu Tyr Lys Tyr Asp His Asn Phe Val Lys
            165             170             175
Ala Ile Asn Ala Ile Gln Lys Ser Trp Thr Ala Thr Thr Tyr Met Glu
            180             185             190
Tyr Glu Thr Leu Thr Leu Gly Asp Met Ile Arg Arg Ser Gly Gly His
        195             200             205
Ser Arg Lys Ile Pro Arg Pro Lys Pro Ala Pro Leu Thr Ala Glu Ile
    210             215             220
Gln Gln Lys Ile Leu His Leu Pro Thr Ser Trp Asp Trp Arg Asn Val
225             230             235                         240
His Gly Ile Asn Phe Val Ser Pro Val Arg Asn Gln Ala Ser Cys Gly
                245             250             255
Ser Cys Tyr Ser Phe Ala Ser Met Gly Met Leu Glu Ala Arg Ile Arg
            260             265             270
Ile Leu Thr Asn Asn Ser Gln Thr Pro Ile Leu Ser Pro Gln Glu Val
        275             280             285
Val Ser Cys Ser Gln Tyr Ala Gln Gly Cys Glu Gly Gly Phe Pro Tyr
    290             295             300
Leu Ile Ala Gly Lys Tyr Ala Gln Asp Phe Gly Leu Val Glu Glu Ala
305             310             315                         320
Cys Phe Pro Tyr Thr Gly Thr Asp Ser Pro Cys Lys Met Lys Glu Asp
            325             330             335
Cys Phe Arg Tyr Tyr Ser Ser Glu Tyr His Tyr Val Gly Gly Phe Tyr
            340             345             350
Gly Gly Cys Asn Glu Ala Leu Met Lys Leu Glu Leu Val His His Gly
        355             360             365
Pro Met Ala Val Ala Phe Glu Val Tyr Asp Asp Phe Leu His Tyr Lys
    370             375             380
Lys Gly Ile Tyr His His Thr Gly Leu Arg Asp Pro Phe Asn Pro Phe
385             390             395                         400
Glu Leu Thr Asn His Ala Val Leu Leu Val Gly Tyr Gly Thr Asp Ser
            405             410             415
Ala Ser Gly Met Asp Tyr Trp Ile Val Lys Asn Ser Trp Gly Thr Gly
            420             425             430
Trp Gly Glu Asn Gly Tyr Phe Arg Ile Arg Arg Gly Thr Asp Glu Cys
        435             440             445
Ala Ile Glu Ser Ile Ala Val Ala Ala Thr Pro Ile Pro Lys Leu
450             455             460
```

```
<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 4
aggaggttgt gtcttgtagc c
```

```
<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 5
agtgcctgtg taggggaagc                                                        20


<210> 6
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 6
tgtcaactgc tcggttatgg gtaa                                                   24


<210> 7
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 7
tcgagcttct cttcgtacac cact                                                   24


<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 8
tgactacaag tggtttgcct tttt                                                   24


<210> 9
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 9
tgctgccctc ttctttatac tgc                                                    23


<210> 10
<211> 24
<212> DNA
<213> Artificial Sequence
```

```
<220>
<223> Description of Artificial Sequence: Primer


<400> 10
gcctctgaga atgtgtatgt caac                                           24


<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 11
cctgccccaa aaatgagata                                                20


<210> 12
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 12
tcgaaaaatc ccaaggtaat c                                              21


<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 13
gggcctagaa aggaaatata catt                                           24


<210> 14
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 14
aatttgttcg gaactattta ttga                                           24


<210> 15
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
```

```
<400> 15
tcgcttctag catacccata                                        20


<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 16
ggcgatcaga ctggcacact agaa                                   24


<210> 17
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 17
cttacccata accgagcagt tgac                                   24


<210> 18
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 18
gcagactgtg ctcaaactgg gtag                                   24


<210> 19
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 19
tctactaatc agaagaggtt tcag                                   24


<210> 20
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 20
ggcacattta ctgtgaatga gag                                    23
```

<210> 21
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 21
gtctcatttg tagcaactca c   21

<210> 22
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 22
ccactttcca cttaggcaca g   21

<210> 23
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 23
aggatggtat tcagcattca ta   22

<210> 24
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 24
atcctagcta gtctggtagc tgaa   24

<210> 25
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 25
tctaggtatc cccgaaatcc atca   24

<210> 26
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 26
gattctctgt gaggcttcag atgt   24

<210> 27
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 27
gccaacaaca gccagctgca caca   24

<210> 28
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 28
tccccactta accactttgc   20

<210> 29
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 29
acttcatagc tgaccatctt cc   22

<210> 30
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 30
gccgccctcc tgctgcttct   20

<210> 31
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 31
ggcttaggat tggggtctga   20

```
<210> 32
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 32
gtatgctaga agcgagaatc cgtat                                             25


<210> 33
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 33
ccaatgctaa aacttgttga gacc                                             24
```